

Pervasive adaptation of gene expression in *Drosophila*

Armita Nourmohammad^{1,2*}, Joachim Rambeau^{2,3}, Torsten Held²,
Johannes Berg², Michael Lässig^{2*}

¹ Joseph-Henri Laboratories of Physics and Lewis-Sigler Institute for Integrative Genomics,
Princeton University, Princeton, NJ 08544, USA

² Institut für Theoretische Physik, Universität zu Köln,
Zùlpicherstr. 77, 50937, Köln, Germany

³ Dynamique des Interactions Membranaires Normales et Pathologiques, UMR5235 CNRS,
Université de Montpellier, 34095 Montpellier, France

Gene expression levels are important molecular quantitative traits that link genotypes to molecular functions and fitness. In *Drosophila*, population-genetic studies in recent years have revealed substantial adaptive evolution at the genomic level. However, the evolutionary modes of gene expression have remained controversial. Here we present evidence that adaptation dominates the evolution of gene expression levels in flies. We show that 63% of the observed expression divergence across seven *Drosophila* species are adaptive changes driven by directional selection. Our results are derived from the variation of expression within species and the time-resolved divergence across a family of related species, using a new inference method for selection. Adaptive gene expression is stronger in specific functional classes, which include regulation, sensory perception, sexual behavior, and morphology. Adaptation increases with broader codon usage and, consistently, highly expressed genes contribute less to adaptation. While most genes show coherent evolution in both sexes, we identify a large group of genes with sex-specific adaptation of expression, which predominantly occurs in males. Our analysis opens a new avenue to map system-wide selection on molecular quantitative traits independently of their genetic basis.

* Correspondence should be addressed to: Armita Nourmohammad (armitan@princeton.edu) or Michael Lässig (mlaessig@uni-koeln.de).

In recent years, several studies have found evidence for widespread adaptive evolution of the *Drosophila* genome [1–3]. This includes adaptive changes in non-coding sequence, consistent with classical ideas on the importance of regulatory evolution for phenotypic adaptation [4]. Gene expression levels are important molecular phenotypes that quantify the effects of regulation on organismic traits and fitness. Insights on how genome evolution affects gene expression have come from studies of quantitative trait loci (QTL); see refs. [5–7] for recent reviews. In yeast, at least 10% of the genes have been inferred to undergo adaptive evolution of expression [8]. In flies, expression-QTL analysis has been used to estimate cis- and trans- effects on expression [9, 10] and to compare the evolution of expression and that of the underlying regulatory sequence [11]; related studies have been performed in yeast [12]. However, given the complexity of the regulatory genotype-to-phenotype map and the limited sensitivity of QTL studies, our understanding of how adaptive genome changes relate to mRNA and protein levels has remained incomplete [5, 7, 13].

An alternative approach is to analyze the evolution of gene expression by methods of quantitative genetics, without explicit reference to genetic evolution of the QTL [6, 7, 14–21]. These studies compare the expression divergence across species, the variation within species, and the expected behavior for neutral evolution [22]. A broad picture of evolutionary constraint on gene expression levels caused by stabilizing selection has emerged in a number of species, including *Drosophila* [6, 14, 15, 17, 20]. A comparative study between human and chimpanzee has produced signatures of predominantly neutral evolution of gene expression [16]. Other studies in primates have identified stabilizing selection, as well as lineage- and tissue-specific directional expression changes [6, 17, 19, 21, 23–25]. However, it has remained difficult to demonstrate that positive selection, as opposed to relaxed stabilizing selection, is the evolutionary cause of expression divergence [5]. Thus, estimating the genome-wide contribution of adaptation to the evolution of gene expression is an outstanding problem.

In this paper, we show that adaptation is the prevalent evolutionary mode of gene expression throughout the *Drosophila* genus. Our method is based on recent theoretical results on the evolution of molecular quantitative traits [26–28] and has two essential features. First, we infer adaptation driven by directional selection together with conservation under stabilizing selection, which allows us to discriminate these two modes in the evolution of gene expression. Second, we identify observables that decouple from number and effects of the underlying QTL. These molecular determinants of gene expression are often unknown, vary considerably between genes, and confound a naive phenotype-based inference of selection.

Results

The pattern of gene expression divergence

We use gene expression data from samples of inbred male and female individuals [19], which cover 6332 orthologous genes in seven *Drosophila* species. A phylogenetic tree of these species is shown in Fig. S1. Gene expression levels are defined in a standard way as logarithms of mRNA counts, suitably normalized to account for differences in assay sensitivity between

experimental probes (Materials and Methods). Within each population, we use these data to estimate the mean expression level of a gene, its total heritable expression variance (referred to as expression diversity), and its non-heritable expression variance between inbred individuals. For each pair of populations, we obtain the cross-species expression divergence of a gene as the squared difference between the population mean levels (see Materials and Methods for details). These quantities define the (suitably scaled) divergence-diversity ratio Ω , which plays a key role in our evolutionary analysis (Box 1 and Materials and Methods). For a given gene, the evolution of the Ω ratio depends only weakly on the effect distribution of its expression QTL and on the amount of recombination between these loci. For neutral evolution, this property is implicitly contained in classical quantitative genetics approaches [22, 29], but it holds more generally under stabilizing and directional selection [26–28].

To obtain a genome-wide evolutionary picture of gene expression in *Drosophila*, we evaluate the aggregate divergence-diversity ratio for all genes in our data set (Materials and Methods). Grouping the species into 6 clades, we obtain a consistent pattern of the Ω ratio as a function of divergence time, τ (Fig. 1A). These data show macro-evolution of expression levels: already the average expression divergence for the youngest clade, *D. melanogaster* and *D. simulans*, is by a factor 2 larger than the average diversity within species (Fig. S1). On the other hand, the observed Ω values for all clades are substantially smaller than the expected values under neutral evolution, which can be computed analytically [26]. This constraint suggests that gene expression evolves under stabilizing selection, in agreement with previous studies [14, 15, 20] and with a standard Q_{ST}/F_{ST} analysis [29] (Materials and Methods). Importantly, however, the Ω ratio increases with divergence time throughout the *Drosophila* genus and does not show evidence of saturation for larger values of τ , in accordance with a similar pattern of the expression divergence observed previously [19]. In the following, we will show that this feature reflects adaptive evolution of gene expression.

Fitness model for gene expression

The inference of adaptation is based on a minimal dynamical model of selection: gene expression levels evolve in a single-peak fitness seascape [26, 28]. This model is illustrated in Box 1 and formally defined in Materials and Methods. The fitness peak for a given gene performs a random walk over macro-evolutionary periods. This walk maps continual changes of the optimal expression of that gene, which are generated by long-term environmental shifts and epistatic co-evolution with other genes. Despite its simplicity, the seascape model combines two salient features of selection on gene expression: stabilizing selection generates evolutionary constraint, and directional selection drives long-term adaptive changes. These selection components are measured by two parameters, the stabilizing strength c and the driving rate v .

Here we use the dependence of the divergence-diversity ratio on evolutionary time, $\Omega(\tau)$, to infer the fitness seascape of gene expression (Box 1). This method discriminates directional selection in a genuine fitness seascape from purely stabilizing selection in a static fitness landscape, providing a more powerful inference of adaptive evolution than Q_{ST}/F_{ST} analy-

sis [29] (Materials and Methods). It also estimates the most important summary statistics of the adaptive process: the cumulative fitness flux Φ , which measures the fitness gain through adaptive expression changes over an evolutionary period (Box 1, Materials and Methods).

The fitness seascape of *Drosophila* gene expression

We first use the aggregate divergence-diversity data to infer a gene-averaged fitness seascape of expression levels in *Drosophila* (Fig. 1A). The least-square fitted seascape model (green line) contains stabilizing and directional selection, leading to adaptive evolution. This model explains the observed pattern $\Omega(\tau)$: the short-term evolutionary constraint is caused by stabilizing selection, and the approximately linear long-term increase signals adaptation. Because genetic drift and adaptation differ in tempo, the contribution of adaptation to expression divergence depends on evolutionary time: the adaptive part is small for the youngest species clades, but adaptation becomes dominant across the entire *Drosophila* genus (green shaded area). In contrast, stabilizing selection alone cannot explain the *Drosophila* expression data. In a static fitness landscape, genetic drift generates a rapidly saturating pattern of $\Omega(\tau)$ that is not observed in the data (Fig. S2).

We can extend the seascape inference to individual genes, using a Bayesian inference scheme that decouples from number and effects of their expression QTL (Materials and Methods). We obtain a posterior distribution of stabilizing strength and fitness flux for each gene. For 54% of all genes, we infer a significant cumulative fitness flux Φ across the *Drosophila* genus; we classify these genes as adaptively regulated (Table S1, Materials and Methods). Fig. 1B shows the average posterior distribution of the fitness flux, which determines a clade-specific adaptive fraction of the expression divergence (Table 1, Materials and Methods). This fraction increases with clade divergence time, in accordance with the aggregate data of Fig. 1A. Between *D. melanogaster* and *D. simulans*, which diverged about 2-3 Myrs ago, 92% of the expression divergence can be attributed to genetic drift under stabilizing selection. Across the entire *Drosophila* genus, which has its last common ancestor about 40 Myrs ago, we infer 63% of the expression divergence to be adaptive. The Bayesian scheme also allows us to quantify the overall statistical significance of our selection inference. In Fig. 1C, we plot the cumulative log-likelihood score for all genes as a function of c and Φ . As shown by a log-likelihood test, the global maximum-likelihood seascape model is strongly favored over the maximum-likelihood landscape model ($P < 10^{-3600}$) and over neutral evolution ($P < 10^{-5400}$) (Materials and Methods). We note that this analysis rejects neutral evolution and evolution under static stabilizing selection independently of model assumptions on the adaptive dynamics.

Testing alternative evolutionary scenarios

The minimal seascape model explains the pattern of gene expression divergence across the *Drosophila* genus in a parsimonious way. But are there equally parsimonious alternative modes of selection or demography that are consistent with the data? To assess the speci-

ficity and robustness of the seascape-based inference, we characterize the statistics of gene expression levels in a number of alternative modes of evolution by analytical approximations and simulations, and we compare the results to the *Drosophila* data.

First, demographic effects may increase or decrease the effective population size in a specific lineage, which affects the stabilizing strength c for all genes. As shown in Fig. S3, lineage-specific changes in effective population size that persist over sufficiently long evolutionary periods can be traced in the aggregate divergence-diversity function $\Omega(\tau)$. Such effects are not observed in our data, which suggests that long-term demographic effects do not play a dominant role in the evolution of *Drosophila* gene expression levels (Fig. S3). This result does not exclude short-term changes of population size, which occurred, for example, in the recent evolution of the *D. melanogaster* lineage [30]. Such changes can be traced in sequence polymorphism spectra [31–34], but they have only minor effects on gene expression levels.

Next, we ask if the *Drosophila* data can be explained by lineage- and gene-specific relaxation of stabilizing selection. We consider a specific non-adaptive mode of expression changes: functional genes evolve under stabilizing selection in a static fitness landscape, but individual genes can (partially) lose function at a given point in their evolutionary history, which relaxes selection on their expression. We model loss of function as stochastic events occurring at a small rate, independently for each gene and on each lineage. This model produces a divergence-diversity function $\Omega(\tau)$ with a long-term nonlinearity that is not seen in the Ω data (Fig. S4). The most direct way to discriminate between relaxation of selection and adaptive evolution is to use a directional bias: most functional genes are up-regulated by stabilizing selection (a similar bias has been exploited in expression QTL studies [5, 8, 25]). In the loss-of-function mode, a comparison of expression levels for a given gene should show small cross-species differences at higher expression levels (i.e., between the lineages with a functional gene) together with large deviations at lower levels (i.e., in the lineages with lost gene function). Accordingly, the distribution of expression divergence values for a given species pair should show a broad tail generated by the loss events. These features are not observed in our data, indicating that relaxed stabilizing selection alone cannot explain the evolution of *Drosophila* expression levels (Fig. S4). Of course, loss of gene function does happen in our phylogeny, but the affected genes will often lose expression altogether and, hence, will be suppressed in our data set.

We can also compare the *Drosophila* data with alternative models of adaptive evolution. For example, individual genes can undergo a (partial) neo-functionalization that requires a major change in their expression. We describe this mode of evolution by a punctuated fitness seascape, in which large shifts of the peak position are stochastic events occurring at a small rate [26]. This process produces an aggregate divergence-diversity function $\Omega(\tau)$ that is compatible with the data, but a broad tail in the distribution of expression divergence values that is not observed (Fig. S4). We conclude that gradual but continual changes in optimal levels, as described by our minimal model, are the dominant evolutionary force driving the adaptation of gene expression in *Drosophila*.

Functional and mechanistic determinants of selection

By applying our inference to specific classes of genes, we can get a more detailed view on adaptation of gene expression in *Drosophila*. First, we observe a strong correlation between codon usage and adaptation: genes with specific codons show strongly reduced adaptive expression divergence and lower average fitness flux than genes with broad codon usage (Fig. 2A,B, Table 1). Specific codon usage is known to be prevalent in highly expressed genes [35]; consistently, we find stronger conservation of expression and lower levels of fitness flux in this class (Table 1). Different codons for the same amino acid differ in their efficiency of translation [35,36], which implies that genes with broad codon usage have a higher potential for adaptive expression changes at the post-transcriptional level. Here we find stronger adaptation at the mRNA level in this gene class, which suggests a two-tier mode of evolution: adaptive mRNA changes lay the ground on which coherent adaptive tuning of protein levels can build.

At the same time, we find no significant correlation between the fitness flux for expression changes and adaptive evolution of amino acid sequence, as measured by a McDonald-Kreitman test [37] (Fig. S5). We conclude that gene expression and gene function provide two largely independent modes of evolution. For a metabolite or a transcription factor, adaptive changes of its cellular concentration are often coupled with conservation of its function.

Our gene-specific inference can be used to detect functional gene classes associated with adaptive evolution of regulation. A full ranking of gene classes by enrichment in adaptively regulated genes with associated P values is reported in Table S1. Gene functions associated with enhanced adaptive evolution of expression include sensory perception, regulation, neural maturation, regulation of growth, aging and morphology. Adaptively regulated functions also include response to UV radiation, which has recently been identified as an important climate-mediated trait in humans [25,38]. Adaptive evolution of genes related to growth, regulation and morphology has been previously inferred by expression QTL and comparative studies of gene regulation in other species [5,6,21]. Here we identify these categories from a quantitative, system-wide scan for adaptively regulated genes. This points to the power of our phenotype-based inference scheme, which is not confounded by the combinatorial complexity of cis-regulatory sequence in higher eukaryotes.

Sex-specific evolution of expression

We can also test the role of expression differentiation between male and female individuals for adaptive evolution across the *Drosophila* genus. The sex specificity of a given gene [19], defined as the difference between its male and female expression level $E_{mf} = E_m - E_f$, is a distinct quantitative trait whose evolutionary pattern can be analyzed by our method. We can distinguish two modes of evolution: conservation of sex specificity maintained by stabilizing selection and sex-specific adaptation of expression (Fig. 2C). Most genes of our data set have well-conserved and often small sex specificity; these genes evolve their expression levels coherently between males and females [19]. The remaining 19% of the genes have a

significant cumulative fitness flux of their specificity trait, Φ_{mf} ; we classify them as undergoing sex-specific adaptation of expression in the *Drosophila* genus. These genes cover all four chromosomes of the *Drosophila* genome.

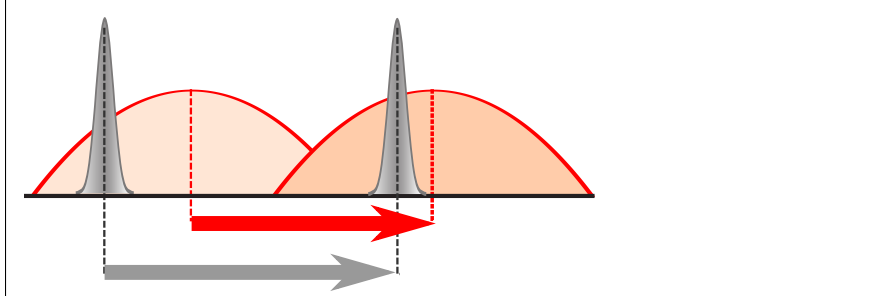
Gene functions associated with sex-specific adaptation of expression include regulation of translation, reproduction, post-mating behavior and (immune) response to biotic stimuli (Table S2). To understand the distribution of these adaptive processes between sexes, we apply our inference to classes of genes with different species-averaged sex bias of expression [39]. For male-biased genes, the divergence-diversity ratio Ω_{mf} signals substantial sex-specific adaptation (Fig. 2D). Consistently, the fitness flux Φ_{mf} is strongly enhanced in genes that are predominantly expressed in males (Fig. 2E). The fitness flux is lower in the other classes, including genes expressed predominantly in females. Together, we find a remarkable evolutionary asymmetry between sexes: male bias in expression is associated with adaptive evolution of expression (orange shaded areas in Fig. 2D,E); female bias in expression is under weaker directional selection, which suggests it primarily reflects conserved physiological differences between male and female organisms. This result complements a previously observed evolutionary asymmetry at the sequence level: genes with male-biased expression show increased amino acid divergence [19]. As suggested by a McDonald-Kreitman test, this increase can be associated with adaptive evolution of gene function (Fig. S5).

Discussion

We have shown that adaptive regulation accounts for most of the macro-evolutionary divergence in gene expression across the *Drosophila* genus. Genes differ considerably in the amount of adaptation, depending on their codon usage, sexual differentiation, and functional class. These results provide evidence for system-wide adaptation of gene regulation in *Drosophila* already at the primary level of transcription, notwithstanding further evolutionary complexities at the level of translation [6, 12]. It remains to be seen whether a similar prevalence of adaptation in the evolution of expression will be found in different species.

Our inference of adaptation is based on the expression divergence-diversity ratio, which depends on the evolutionary distance between species. It exploits two fundamental evolutionary features of quantitative traits: at short evolutionary distance, the divergence is always near neutrality; at longer distance, it depends jointly on stabilizing and directional selection. These features generate a divergence pattern with two distinct molecular clocks, as shown in Fig. 1. Importantly, the phenotypic evolution of gene expression decouples from details of its genetic basis. This explains why we find strong selection on gene expression levels although selection on individual QTL is often weak [40].

Our method can be applied to a broad spectrum of molecular quantitative traits with a complex genetic basis, provided comparative data from multiple, sufficiently diverged species are available. Such traits include genome-wide protein levels, protein-DNA binding interactions or enzymatic activities. For most of these traits, we have only partial knowledge of the underlying genetic loci and their effects on trait and fitness. Our method complements QTL studies and opens a way to infer quantitative phenotype-fitness maps at the systems level.



Box 1: Inferring adaptive evolution of quantitative traits.

The schematic shows the evolution of a quantitative trait in a single-peak fitness seascape. The distribution of trait values within a species (grey curves) changes over macro-evolutionary periods, which can be observed as cross-species divergence of the mean trait values (grey arrow). In the underlying fitness seascape (red curves), evolutionary displacements of the fitness peak lead to lineage-specific optimal trait values and directional selection (red arrow). The minimal fitness model has two parameters: the stabilizing strength c is proportional to the inverse square width of the fitness peak, the driving rate v measures the mean square displacement of the fitness peak per unit of evolutionary time (see Materials and Methods for precise definitions). We infer selection on quantitative traits from their time-resolved inter-species divergence, $D(\tau)$, and their intra-population diversity, Δ , as defined in Materials and Methods. We evaluate the ratio $\Omega(\tau) = \pi_0 D(\tau) / \Delta$, scaled by the neutral sequence diversity π_0 and averaged over a family of traits for three or more species with different divergence times τ (Fig. 1A, squares). The Ω test [26] is guided by theoretical results on the evolution of quantitative traits [26, 28]. These provide the analytical form of $\Omega(\tau)$ in a fitness seascape ($c > 0$, $v > 0$; green solid line). The corresponding form $\Omega_{\text{eq}}(\tau)$ in a fitness landscape of the same stabilizing strength ($c > 0$, $v = 0$; blue solid line) reaches a saturation value Ω_{stab} . Fitting the seascape model to the Ω data determines the decomposition $\Omega(\tau) = \Omega_{\text{eq}}(\tau) + \Omega_{\text{ad}}(\tau)$ (blue and green shaded areas). The amplitude ratio $\omega_{\text{ad}}(\tau) = \Omega_{\text{ad}}(\tau) / \Omega(\tau)$ gives the fraction of trait divergence that is adaptive, i.e., driven by directional selection. In the linear regime $\Omega(\tau) \approx \Omega_{\text{stab}} + \Omega_{\text{ad}}(\tau)$, which covers all species clades in this dataset, the fitted amplitudes provide simple estimates of the selection parameters, $c \approx 1 / (2\Omega_{\text{stab}})$ and $v \approx 2\Omega_{\text{ad}}(\tau) / \tau$, and of the resulting cumulative fitness flux $2N\Phi = 2cv\tau$ (scaled by the effective population size N). The divergence-diversity ratio for neutral evolution, $\Omega_0(\tau)$, is shown for reference ($c = 0$; grey solid line). A Bayesian extension of the Ω test and its relationship to other trait- and sequence-based selection tests ($Q_{\text{st}}/F_{\text{st}}$ test [29], Ornstein-Uhlenbeck models, [20, 43–45] and McDonald-Kreitman test [37]) are discussed in Materials and Methods.

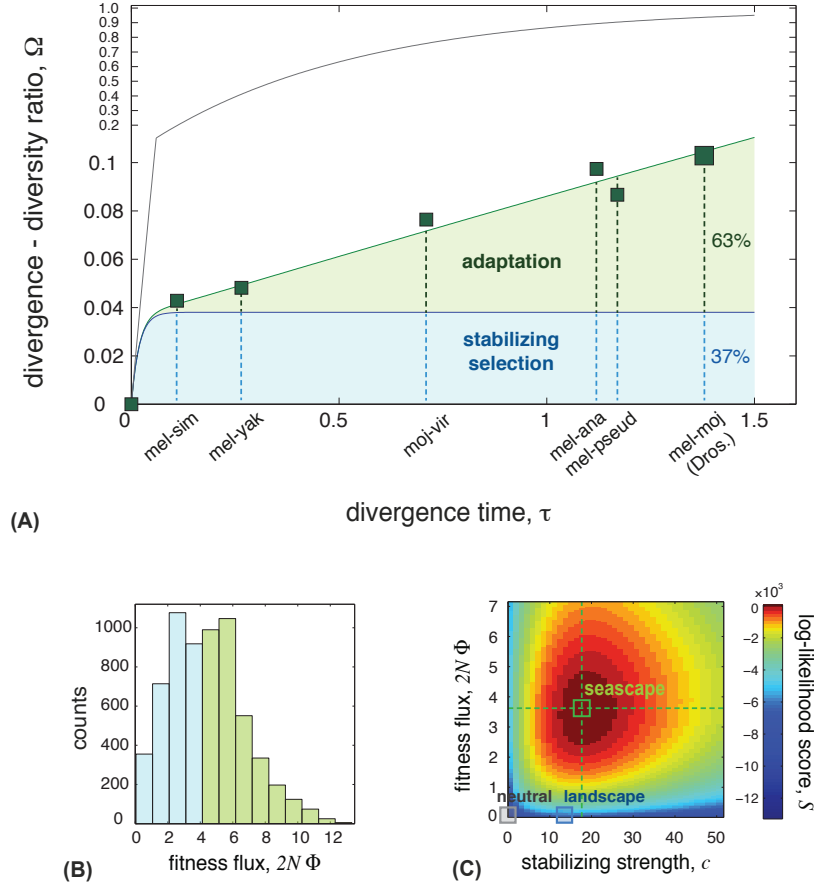


Figure 1: Adaptive evolution of gene expression. (A) The aggregate divergence-diversity ratio, Ω , from all genes of this data set is plotted against the divergence time, τ for six partial species clades (■) and for the entire *Drosophila* genus (■). Species clades and divergence times (scaled by the rate of synonymous mutations) are defined by the phylogenetic tree of the *Drosophila* Reference Panel (Fig. S1A and Materials and Methods). These data are shown together with theoretical curves $\Omega(\tau)$ under directional selection (green line), under stabilizing selection (blue line), and for neutral evolution (grey line). Inferred model parameters are $c^* = 18.6$ and $v^* = 0.07$; see Box 1 and Materials and Methods for model details. We infer a time-dependent adaptive component of the expression divergence (green shaded area); the complementary component is generated by genetic drift under stabilizing selection (blue shaded area). Adaptation accounts for 63% of the expression divergence across the *Drosophila* genus. See Fig. S2 for a comparison of the same data to models of time-independent stabilizing selection [20]. (B) Distribution of maximum-likelihood values of the scaled cumulative fitness flux, $2N\Phi$, inferred for individual genes (Materials and Methods). Our inference classifies 54% of all genes as adaptively regulated ($2N\Phi > 4$, green shaded part of the distribution). (C) Bayesian inference of fitness models. The posterior log-likelihood score $S(c, \Phi)$ favors the optimal seascape model ($c^* = 18.6, 2N\Phi^* = 3.6$; green square) over the best landscape model ($c_{eq}^* = 16, \Phi = 0$; blue square) and the neutral model ($c = 0, \Phi = 0$; grey square).

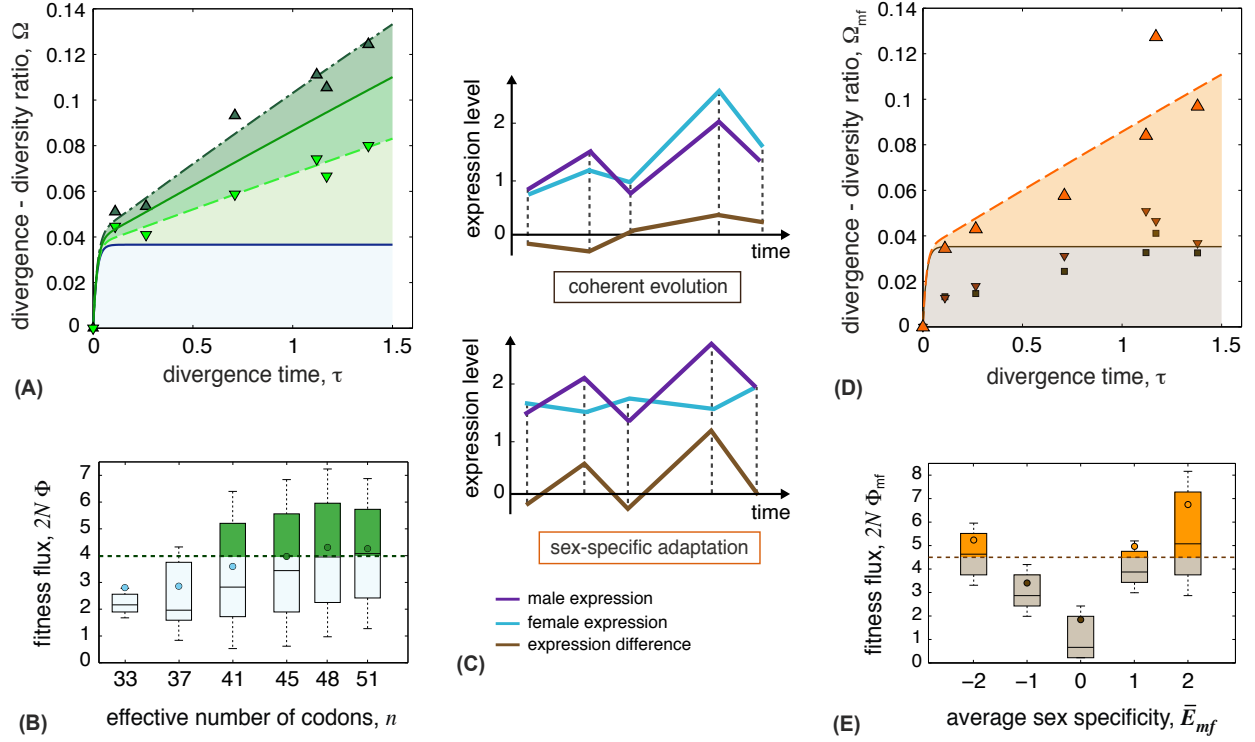


Figure 2: Adaptation depends on codon usage and sex bias. (A) The aggregate divergence- diversity ratio, Ω , for genes with broad codon usage (Δ) and for genes with specific codon usage (∇) is shown together with theoretical curves under directional selection (dashed and dashed-dotted lines); the theoretical curve inferred for all genes is shown for comparison (solid line, cf. Fig. 1A). Codon usage is measured by the effective number of codons [46] (Materials and Methods); inferred model parameters are listed in Table 1. (B) The distribution of the cumulative fitness flux, $2N\Phi$, is plotted against the effective number of codons [46], n (Materials and Methods; circle: average, line: median, box: 50% around median, bars: 70% around median). (C) Conservation of sex specificity (top panel) vs. sex-specific adaptation of expression (bottom panel). The sex-specificity trait (brown line) is defined as the difference between male and female expression levels (purple and blue lines). The schematics show all three lines as functions of evolutionary time. (D) The aggregate divergence- diversity ratio of the sex-specificity trait, Ω_{mf} , for all genes (\square), genes with male-biased expression (Δ), and genes with female-biased expression (∇), is shown together with theoretical curves under directional selection. (E) The distribution of the cumulative fitness flux for the sex-specificity trait, $2N\Phi_{mf}$, is plotted against the species-averaged sex specificity, \bar{E}_{mf} (Materials and Methods). Sex-specific adaptation ($2N\Phi_{mf} > 4.5$, orange shaded part) occurs predominantly in male-biased genes.

Materials and Methods

Summary

Synonymous genome sequence is used to estimate the neutral sequence diversity $\pi_0 = 4\mu N$ (ref. [41], μ is the point mutation rate and N the effective population size) and the species divergence times τ_{ij} (ref. [42], scaled in units of μ^{-1}); these data underlie the phylogenetic tree [42] (Fig. S1). The expression levels $E_{i,s,\kappa}^\alpha$ measured by Zhang et al. [19] are labelled by gene α , species i , sex s , and replicates κ of an inbred line. The levels are normalized to mean 0 and cross-gene variance 1 for each individual; the effects of this normalization on our analysis is tested in Fig. S6. For a given gene α , we estimate isogenic variance between experimental replicates of an inbred line δ_i^α and genetic mean Γ_i^α in each species, and the divergence $D_{ij}^\alpha = (\Gamma_i^\alpha - \Gamma_j^\alpha)^2$ between any two species i, j ; we also estimate genetic variance (expression diversity) Δ^α using data from the two distinct strains of *D. simulans*. We define the aggregate divergence-diversity ratio $\Omega_{ij}^\alpha = \pi_0 \langle D_{ij} \rangle / \langle \Delta \rangle$, where angular brackets denote averages over genes. For each clade \mathcal{C} in our phylogeny, we obtain the aggregate data $(\tau_{\mathcal{C}}, \Omega_{\mathcal{C}})$ shown in Figs. 1A, 2A by averaging τ_{ij} and Ω_{ij} over all pairs of species $i, j \in \mathcal{C}$ that are connected via the root of the clade.

The minimal fitness seascape for a given gene takes the form

$$f(E, t) = -\frac{c}{2NE_0^2} (E - E^*(t))^2,$$

where the optimal trait value $E^*(t)$ performs an Ornstein-Uhlenbeck process with mean square displacement vE_0^2 per unit μ^{-1} of evolutionary time; E_0^2 is the average genetic variation of expression in the long-term limit of neutral evolution [27]. This non-equilibrium model generates adaptive evolution with an average scaled cumulative fitness flux $\langle 2N\Phi \rangle = 2cv\tau_{\text{Dros.}}$ across the *Drosophila* genus ($\tau_{\text{Dros.}} = 1.4$; Fig. S1). Applying the Ω test (Box 1) to aggregate expression data $(\tau_{\mathcal{C}}, \Omega_{\mathcal{C}})$, we infer a global fitness seascape with parameters ($c^* = 18.6, v^* = 0.07$) and an average fitness flux $2N\Phi^* = 3.6$ per gene. Control fits of the same data to equilibrium models, including the well-known Ornstein-Uhlenbeck dynamics for the population mean trait [20, 43–45], are shown in Fig. S2. The probabilistic extension of this test evaluates the Bayesian posterior probability distribution $Q(c, \Phi | \mathbf{E}^\alpha)$ of individual genes, given their sample mean data $\mathbf{E}^\alpha = (E_1^\alpha, \dots, E_T^\alpha)$. This produces gene-specific expectation values c^α and Φ^α (Figs. 1B and 2B); we use the condition $2N\Phi^\alpha > 4$ to infer adaptively regulated genes (Table S1). The cumulative log-likelihood score $S(c, \Phi) = \sum_\alpha \log Q(c, \Phi | \mathbf{E}^\alpha)$ serves to quantify the statistical significance of our inference (Fig. 1C).

To test for lineage-specific demographic effects, we compare the aggregate Ω data to theoretical functions $\Omega(\tau, \tau_i)$ computed for an alternative model with a change in effective population size on the phylogenetic branch of species i (Fig. S3). We also examine two alternative selection scenarios: relaxed stabilizing selection by partial loss of function (c^α switches to a reduced value with rate γ) and punctuated fitness peak shifts ($E^{*\alpha}$ jumps by an amount of order E_0 with a rate of order $v\mu$) (Fig. S4). The observed distributions of cross-species expression differences are consistent with the minimal seascape model but at variance with both alternative models (Fig. S4).

To infer sex-specific evolution, we define specificity traits as differences between male and female expression levels, $E_{\text{mf},i}^\alpha = E_{m,i}^\alpha - E_{f,i}^\alpha$, for each gene [19]. Genes with sex-specific adaptive evolution of expression are identified by a condition on the cumulative fitness flux for the specificity trait, $2N\Phi_{\text{mf}}^\alpha > 4.5$ (Table S2). Genes with male- and female-biased expression are identified using the

results of ref. [39].

We simulate Fisher-Wright evolution in fitness land- and seascapes to validate our probabilistic inference scheme and to establish its robustness under trait epistasis (Fig. S7).

1. Data and primary analysis

Sequence data and phylogeny. Our inference procedure requires the following global sequence-based information (which does not include expression QTL):

- (a) A phylogenetic tree of the 7 *Drosophila* species included in this study. Here we use the tree of the *Drosophila* 12 Genome Consortium [42], which is based on genome-wide divergence at synonymous sequence sites. This tree determines six clades of phylogenetically related species (Fig. S1A), which are used in our analysis of time-dependent expression divergence (Figs. 1A and 2A,B).
- (b) Divergence times between all pairs of species, scaled in units of the inverse neutral point mutation rate. The tree of Fig. S1A uses a lineage-specific mutation rate to infer the length of its 12 branches. The scaled divergence time τ_{ij} for a given species pair (i, j) is the sum of the lengths of the branches connecting these species. The scaled divergence time of a clade \mathcal{C} is defined as an average over species pairs,

$$\tau_{\mathcal{C}} = \frac{1}{|\mathcal{C}_1||\mathcal{C} \setminus \mathcal{C}_1|} \sum_{i \in \mathcal{C}_1} \sum_{j \in \mathcal{C} \setminus \mathcal{C}_1} \tau_{ij}, \quad (1)$$

where \mathcal{C} is the set of species in the clade and $(\mathcal{C}_1, \mathcal{C}_2)$ is the portioning of this set defined by the root node. These divergence times differ substantially from previous estimates based on amino acid distances [20].

- (c) The neutral nucleotide sequence diversity,

$$\pi_0 = 2p\mu N, \quad (2)$$

where N is the effective population size, and $p = 1, 2$ for haploid/diploid organisms, respectively. Here we use $\pi_0 = 0.0112$, as determined in ref. [41] from genome-wide polymorphism data at synonymous sequence sites. The sequence diversity enters the definition of the scaled Ω ratio in equation 9 and the probabilistic extension of the Ω test (section 2).

Expression data. We use genome-wide expression data from 7 *Drosophila* species obtained by Zhang et al. [19] (Gene Expression Omnibus under accession number GSE6640). These data are well suited for our analysis. They cover several clades of species that are well comparable at the organismic level and sufficiently diverged for adaptive evolution of expression to be detectable (section 2). Moreover, *Drosophila* has larger effective population size,

higher mutation rates, and shorter generation times than typical mammalian species [47], and adaptive evolution has been detected at the genomic level by several methods [1–3]. Hence, compared to more recent data from other species [24, 48, 49], the *Drosophila* expression data of Zhang et al. [19] are a suitable target for the inference of adaptive evolution. These data contain mRNA intensity measurements for a number of male and female inbred replicates in each species. Specific microarray platforms were designed for each of these species. Each platform has an array of probes mapped to assembled genome sequences and to GLEANR gene annotations by the *Drosophila* 12 Genomes Consortium [42], which also provides sequence homology tables. We restrict the analysis to the 6332 genes that have unambiguous one-to-one orthologs across all lines and are tested by at least four probes in each microarray platform. We obtain a set of expression levels $E_{i,s,\kappa}^\alpha$ (defined as \log_2 intensities) labelled by gene number $\alpha \in \{1, \dots, g=6332\}$, species $i \in \{mel, sim, yak, ana, pse, vir, moj\}$ (Fig. S1), sex $s \in \{m, f\}$, and inbred replicates $\kappa \in \{1, \dots, k_{i,s} = 4 - 7\}$. The data contain two strains of *D. simulans* (14021-0251.011 and 14021-0251.198), which are used to estimate the genetic variance of expression (see below).

Normalization of expression levels. We define a linear transformation of the levels [50],

$$E_{i,s,\kappa}^\alpha \rightarrow \frac{E_{i,s,\kappa}^\alpha - \langle E_{i,s,\kappa} \rangle}{\sqrt{V_{i,s,\kappa}}}, \quad (3)$$

where $\langle E_{i,s,\kappa} \rangle$ and $V_{i,s,\kappa}$ denote mean and variance of the expression across all genes in a given individual (i, s, κ) . The transformed levels $E_{i,s,\kappa}^\alpha$ are shifted to mean 0 and normalized to variance 1 across all genes in each individual. The transformation (3) is a heuristic to reduce differences in probe sensitivity between microarrays (each individual is measured in a separate array). To test its influence on our inference of selection, we compare the aggregate Ω ratio, which is defined in equation (9), for untransformed expression levels, expression levels with only shift ($E_{i,s,\kappa}^\alpha \rightarrow E_{i,s,\kappa}^\alpha - \langle E_{i,s,\kappa} \rangle$), and expression levels with shift and normalization [50] (Fig. S6A-C). Shifting to zero mean turns out to be an essential step to remove spurious expression divergence. The subsequent normalization to variance 1 affects the Ω data and our inference of selection only weakly. However, additively and multiplicatively transformed expression levels produce less noisy Ω data than levels with only additive transformation (cf. section 2). Hence, we use levels $E_{i,s,\kappa}^\alpha$ as given by the transformation (3) for our evolutionary analysis.

Expression statistics within and across populations. Using the normalized expression levels, we can define averages and natural variation of expression at three different levels:

- (a) The mean and variance of expression across experimental replicates of an inbred line characterize the distribution of expression levels for a given genotype. Here we estimate

these quantities from the data of each inbred line,

$$E_{i,s}^\alpha = \frac{1}{k_i} \sum_{\kappa} E_{i,s,\kappa}^\alpha, \quad \delta_{i,s}^\alpha = \frac{1}{k_i - 1} \sum_{\kappa} (E_{i,s,\kappa}^\alpha - E_{i,s}^\alpha)^2, \quad (4)$$

and we define the sample mean and variance,

$$E_i^\alpha = \frac{1}{2}(E_{i,m}^\alpha + E_{i,f}^\alpha), \quad \delta_i^\alpha = \frac{1}{2}(\delta_{i,m}^\alpha + \delta_{i,f}^\alpha). \quad (5)$$

- (b) The genetic mean and diversity of expression characterize the distribution of heritable expression differences in a population. Heritable components of quantitative traits are often inferred from “common garden” breeding experiments under standardized environmental conditions. Here we estimate the genetic mean and diversity for a given gene from the data within one species,

$$\Gamma_i^\alpha = E_i^\alpha \pm \sqrt{\frac{\Delta_i^\alpha}{2} + \frac{\delta_i^\alpha}{k_i}}, \quad \Delta^\alpha = \frac{1}{2}(E_{sim1}^\alpha - E_{sim2}^\alpha)^2 - \frac{1}{k_i} \delta_i^\alpha, \quad (6)$$

where we have included the expected sampling error for Γ_i^α . The data set of ref. [19] limits the inference of expression diversity to a broad estimate from two strains of *D. simulans*. This is sufficient for our analysis, because the inference of adaptation decouples from the precise value of $\langle \Delta \rangle$ (section 3). Similarly, we define the expression dimorphism between males and females in each species,

$$\Delta_{i,mf}^\alpha = \frac{1}{2}(E_{i,m}^\alpha - E_{i,f}^\alpha)^2 - \frac{1}{k_i} \delta_i^\alpha. \quad (7)$$

- (c) The expression divergence is defined as the squared difference between population means, $D_{ij}^\alpha = (\Gamma_i^\alpha - \Gamma_j^\alpha)^2$, and characterizes evolutionary expression differences between two species. Here we estimate the divergence for a given gene from the cross-species data,

$$D_{ij}^\alpha = (E_i^\alpha - E_j^\alpha)^2 - \Delta^\alpha - \frac{1}{k_i} \delta_i^\alpha - \frac{1}{k_j} \delta_j^\alpha. \quad (8)$$

Equations (6) and (8) follow Wright’s decomposition of the variance of a quantitative trait into intra- and inter-species components [59], which underlies the quantitative genetics summary statistics F_{ST} and Q_{ST} (see section 2). For the analysis of sex-specific evolution (section 3), we use the same rationale for the sex-specificity traits $E_{i,mf}^\alpha = E_{i,m}^\alpha - E_{i,f}^\alpha$.

In Fig. S1B, we compare gene-averaged values of isogenic variance, diversity, dimorphism and divergence (these averages are denoted by angular brackets), as well as the cross-gene variance of expression. We find a clear ranking $\langle \delta_i \rangle < \langle \Delta \rangle \lesssim \langle \Delta_{i,mf} \rangle < \langle D_{ij} \rangle < V_i$ for all species i and j , where $V_i = \langle \Gamma_i^2 \rangle \approx 1$ by our normalization. In the Ω test for selection on gene expression, we use diversity and divergence estimates given by equations (6) and (8)

in aggregate measures across groups of species and classes of genes. However, our data set has a low number of individuals per species. Hence, single-gene estimates of diversity and divergence are noisy, which calls for a probabilistic inference of selection. The Ω test and its probabilistic extension for individual genes are described in section 2.

Divergence-diversity ratio, Ω . The aggregate expression divergence-diversity ratio Ω_{ij} for a given species pair (i, j) is defined as

$$\Omega_{ij} = \pi_0 \frac{\langle D_{ij} \rangle}{\langle \Delta \rangle}, \quad (9)$$

where π_0 is the neutral sequence diversity (equation 2). The expression diversity and gene-specific divergence values D_{ij}^α and Δ_{ij}^α are given by equations (6) and (8). Angular brackets denote averages over all genes in our dataset, $\langle D_{ij} \rangle = \frac{1}{g} \sum_\alpha D_{ij}^\alpha$. The prefactor in equation (9) is chosen such that $\Omega \simeq 1$ for neutral evolution in the limit of long divergence times (section 2). The divergence-diversity ratio $\Omega_{\mathcal{C}}$ for a species clade \mathcal{C} is defined as an average over species pairs,

$$\Omega_{\mathcal{C}} = \frac{1}{|\mathcal{C}_1||\mathcal{C} \setminus \mathcal{C}_1|} \sum_{i \in \mathcal{C}_1} \sum_{j \in \mathcal{C} \setminus \mathcal{C}_1} \Omega_{ij}, \quad (10)$$

in analogy with the definition (1) of clade divergence times. We also define divergence-diversity ratios $\Omega_{ij}^{\mathcal{G}}$ and $\Omega_{\mathcal{C}}^{\mathcal{G}}$ for specific gene classes \mathcal{G} , using restricted averages $\langle \dots \rangle_{\mathcal{G}}$.

2. Inference of selection on gene expression

Evolutionary model. We consider the evolution of gene expression levels under genetic drift, mutation, and selection given by a fitness model with peak displacements on macro-evolutionary time scales. In the minimal seascape model [26, 28], the fitness of a given gene depends on its expression level E and on evolutionary time t ,

$$f(E, t) = f^* - c_0 (E - E^*(t))^2. \quad (11)$$

The expression value of maximum fitness, $E^*(t)$, performs an Ornstein-Uhlenbeck random walk with diffusion constant v_0 , average value \mathcal{E} and stationary mean square deviation $r^2 E_0^2$, where r^2 is a constant of order 1. This process is defined by the Langevin equation

$$\frac{d}{dt} E^*(t) = -\frac{v_0}{2r^2 E_0^2} (E^*(t) - \mathcal{E}) + \eta(t), \quad (12)$$

where $\eta(t)$ is the random variable of a delta-correlated Gaussian process with average 0 and variance v_0 . These random variables are assumed to be independent for each gene and on each lineage. The Ornstein-Uhlenbeck fitness seascape should not be confused with a previous Ornstein-Uhlenbeck model for the evolution of quantitative traits under stabilizing selection [20, 43–45, 51–53] (a detailed comparison is given below).

The minimal seascape model captures two kinds of selection on gene expression in a unified way:

- (a) Stabilizing selection. This type of selection constrains the intra- and inter-population variation of expression levels to values around $E^*(t)$. We define the dimensionless *stabilizing strength*

$$c = 2N E_0^2 c_0, \quad (13)$$

where N is the effective population size and the trait scale E_0^2 is given by the average genetic variation of expression in the long-term limit of neutral evolution [27], $E_0^2 = \lim_{\tau \rightarrow \infty} \langle D(\tau) \rangle / 2$. In the limit case $v_0 = 0$, the fitness seascape reduces to a static fitness landscape, $f(E) = f^* - c_0(E - E^*)^2$, and stabilizing selection is the only selective force. This provides a simple interpretation of the selection parameter c : it compares the (hypothetical) genetic load $c_0 E_0^2$ of a neutrally evolving trait evaluated in the landscape $f(E)$ and the actual genetic load $1/2N$ in the same landscape, assuming a mutation-selection-drift equilibrium at low mutation rates [28]. This parameter signals the regimes of weak ($c \lesssim 1$) and strong ($c \gtrsim 1$) stabilizing selection [27].

- (b) Directional selection. In a fitness seascape, this type of selection triggers adaptive response of the population mean trait in the direction of fitness peak displacements. We define the scaled *driving rate*

$$v = \frac{v_0}{\mu E_0^2}. \quad (14)$$

This parameter measures mean square displacement of the fitness peak, in units of E_0^2 and per unit $1/\mu$ of evolutionary time. In *macro-evolutionary* seascapes, v is sufficiently low for population to follow fitness peak displacements; such seascapes are a joint model of stabilizing and directional selection [26]. The values of v inferred from our data fall in this regime (see section 2). Because the seascape dynamics is a short-range Markov process, the mean square peak displacement over a scaled evolutionary time τ is then simply $E_0^2 v \tau$. (Here we express v in units of μ and τ in units of $1/\mu$, which differs slightly from the notation in refs. [26, 28].) In the long-term regime $v \tau \gg r^2$, the fitness peak dynamics becomes stationary with mean \mathcal{E} and variance $r^2 E_0^2$. This regime turns out to be well beyond the divergence times in our species sample. Hence, the statistics of *Drosophila* gene expression levels and our inference of selection are independent of r^2 .

Fitness flux. This measure of adaptation is defined as the speed of movement on a fitness land- or seascape by genotype or heritable phenotype changes in a population [26, 54]. The cumulative fitness flux associated with the population mean expression level $\Gamma(t)$ of a gene in a fitness seascape $f(E, t)$ is given by

$$\Phi(\tau) = \int_{t=0}^{\tau} \frac{\partial f(\Gamma, t)}{\partial \Gamma} \frac{d\Gamma(t)}{dt} dt. \quad (15)$$

This quantity measures the total amount of adaptation over a macro-evolutionary period τ in a population history. This quantity satisfies the fitness flux theorem [54], which generalizes the Fisher’s fundamental theorem of natural selection to mutation-selection-drift processes. As shown by the fitness flux theorem, the average cumulative fitness flux over parallel evolutionary histories, in units of $1/2N$, measures the importance of adaptation compared to genetic drift: adaptation is substantial if $\langle 2N\Phi(\tau) \rangle \gtrsim 1$. For a stationary adaptive process in the minimal seascape (11), the average scaled cumulative fitness flux takes the simple form [26, 28]

$$\langle 2N\Phi(\tau) \rangle \simeq 2cv\tau, \quad (16)$$

up to factors of order π_0 , as quoted in Box 1. The exact functional form of the fitness flux is given in reference [26]. This quantity is closely related to the time-dependent fraction of expression divergence that is adaptive, $\omega_{\text{ad}}(\tau)$ (equation 24). We introduce the shorthand $\Phi = \Phi(\tau_{\text{Dros.}})$ with $\tau_{\text{Dros.}} = 1.4$ (Fig. S1A); this quantity measures the amount of adaptation across the *Drosophila* genus. By the probabilistic inference method discussed below, we obtain expectation values $2N\Phi^\alpha$ of the rescaled fitness flux for individual genes over the divergence time of the *Drosophila* genus (equation 33). We use these values to describe the overall statistics of expression adaptation (Fig. 1B), to infer differences in adaptation between gene classes (Fig. 2B,D; Table 1), and to define significantly adaptive genes (using a threshold $2N\Phi^\alpha > 4$; Table S1). For the analysis of sex-specific adaptation (Fig. 2C,D), we define an analogous fitness flux $2N\Phi_{\text{mf}}$ for sex-specificity traits (section 4).

Evolutionary modes of quantitative traits. In the minimal seascape model, the aggregate Ω ratio defined by equation (9) depends on the divergence time τ and on the selection parameters c and v . We can use this dependence to distinguish three modes of evolution [26, 28]:

- (a) Neutral evolution ($c = 0$). The divergence-diversity ratio has an initially linear increase due to mutations and genetic drift, and it approaches a maximum value 1 with a scaled relaxation time of 1,

$$\Omega_0(\tau) \simeq \begin{cases} \tau & \text{for } \tau \ll 1, \\ 1 & \text{for } \tau \gg 1. \end{cases} \quad (17)$$

The short-term behavior reflects the linear growth of the average divergence [22, 55, 56],

$$\langle D(\tau) \rangle_0 \simeq \frac{\langle \Delta \rangle_0}{\pi_0} \tau \quad \text{for } \tau \ll 1. \quad (18)$$

The growth rate is the average diversity at neutrality divided by the sequence diversity (equation 2). This ratio equals the mutational variance of a quantitative trait as defined in refs. [22, 55, 56], up to a rescaling of evolutionary time to units of $1/\mu$.

- (b) Evolution under stabilizing selection ($c \gtrsim 1, v = 0$). In a static fitness landscape, the divergence-diversity ratio approaches a smaller maximum value, $\Omega_{\text{stab}}(c) < 1$, with a

proportionally shorter relaxation time [26],

$$\Omega_{\text{eq}}(\tau) \simeq \begin{cases} \tau & \text{for } \tau \ll \Omega_{\text{stab}}(c) \\ \Omega_{\text{stab}}(c) & \text{for } \tau \gg \Omega_{\text{stab}}(c). \end{cases} \quad (19)$$

Over a wide range of evolutionary parameters, which includes the inferred values for the data set of this study, the maximum value depends on the stabilizing strength in a simple way, $\Omega_{\text{stab}}(c) \sim 1/(2c)$, with corrections for weaker selection and for larger nucleotide sequence diversity [27].

- (c) Adaptive evolution under stabilizing and directional selection ($c \gtrsim 1, v > 0$). In a genuine fitness seascape, the divergence-diversity ratio acquires an adaptive component,

$$\Omega(\tau) = \Omega_{\text{eq}}(\tau) + \Omega_{\text{ad}}(\tau) = \begin{cases} \tau & \text{for } \tau \ll \Omega_{\text{stab}}(c) \\ \Omega_{\text{stab}}(c) + \frac{1}{2}v[\tau - 2\Omega_{\text{stab}}(c)], & \text{for } \tau \gg \Omega_{\text{stab}}(c), \end{cases} \quad (20)$$

with corrections for τ approaching the saturation time of fitness peak displacements, r^2/v . The universal short-term behavior $\Omega(\tau) \simeq \tau$ (equations 19 and 20) reflects the quasi-neutral growth of the divergence [26, 28],

$$\langle D(\tau) \rangle \simeq \frac{\langle \Delta \rangle(c)}{\pi_0} \tau \quad \text{for } \tau \ll \Omega_{\text{stab}}(c), \quad (21)$$

where $\langle \Delta \rangle(c)$ is the average diversity under selection. As shown by comparison with the neutral behavior (equation 18), selection enters only via the constraint on $\langle \Delta \rangle$. Over a wide range of the stabilizing strength c , this constraint remains weak and $D(\tau)$ evolves near neutrality [27], as long as $\tau \ll \Omega_{\text{stab}}(c)$.

The full analytical form of the functions $\Omega_0(\tau)$ (equation 17), $\Omega_{\text{eq}}(\tau)$ (equation 19), and $\Omega(\tau)$ (equation 20) is given in refs. [26, 28].

The Ω test for selection. The evolutionary statistics of the Ω ratio provides a joint test for stabilizing and directional selection on quantitative traits. We can infer the selection parameters of a seascape model by fitting the function $\Omega(\tau)$ (equation 20) to data (τ, Ω) . This method has the following properties:

- (a) The Ω test requires data (τ, Ω) from species with different divergence times in the regime $\tau \gtrsim \Omega_{\text{stab}}$. In the quasi-neutral regime $\tau \lesssim \Omega_{\text{stab}}$, the Ω ratio is insensitive to selection (equations 17–20).
- (b) By the decomposition (equation 20), the Ω test infers a time-dependent fraction

$$\omega_{\text{ad}}(\tau) = \frac{\Omega_{\text{ad}}(\tau)}{\Omega(\tau)} \quad (22)$$

of the aggregate trait divergence to be adaptive. The complementary fraction, $1 - \omega_{\text{ad}}(\tau)$, is attributed to genetic drift under stabilizing selection.

- (c) We can approximate the divergence-diversity ratio (equation 20) by the linear form $\Omega(\tau) \approx \Omega_{\text{stab}} + \Omega_{\text{ad}}(\tau) = \Omega_{\text{stab}} + v\tau/2$. Therefore, already a linear fit to data produces simple estimates of stabilizing strength and driving rate,

$$c \approx \frac{1}{2\Omega_{\text{stab}}}, \quad v \approx \frac{2\Omega_{\text{ad}}(\tau)}{\tau}, \quad (23)$$

and infers the adaptive fraction of expression divergence, which is related to the average scaled fitness flux [26] (equation 16),

$$\omega_{\text{ad}}(\tau) \approx \frac{\Omega(\tau) - \Omega_{\text{stab}}}{\Omega(\tau)}, \quad \langle 2N\Phi(\tau) \rangle \approx \frac{2\omega_{\text{ad}}(\tau)}{1 - \omega_{\text{ad}}(\tau)} \approx \frac{2(\Omega(\tau) - \Omega_{\text{stab}})}{\Omega_{\text{stab}}}. \quad (24)$$

- (d) The Ω ratio of a quantitative trait depends on the selection parameters c and v , but it decouples from the genetic basis of the trait. Specifically, it depends only weakly on the number and effect size of the underlying QTL [26, 27], on the amount of recombination between these sites [26, 27], and on the nonlinearities in the genotype-phenotype map (trait epistasis; see section 5 and Fig. S7B). The Ω statistics also decouples from details of the selection dynamics; it can be applied to *punctuated* adaptive processes, which have fewer and larger peak displacements [26] (section 3).

Application of the Ω test to gene expression data. In Fig. 1, we compare the model function $\Omega(\tau)$ (equation 20) to aggregate expression data (τ_c, Ω_c) for six *Drosophila* species clades (equations 1 and 10). The best-fit seascape model ($c^* = 18.6$, $v^* = 0.07$; green line) explains these data, which produces evidence for adaptive evolution of gene expression. Using the decomposition into adaptive and drift components (green and blue shaded areas), we obtain a cumulative fitness flux $\langle 2N\Phi \rangle = 3.6$ across the entire *Drosophila* genus (equations 23 and 24). Importantly, the inference of adaptive evolution decouples from the precise overall scale of Ω , which is influenced by our limited information on expression diversity (section 2).

Control analysis of equilibrium models. We can also compare the aggregate expression data (τ_c, Ω_c) to models of time-independent stabilizing selection:

- (a) Fitness landscape model. In contrast to the seascape model, the best-fit landscape model ($c_{\text{eq}} = 13$, $v = 0$) provides a poor fit to the data (Fig. S2A). It captures the average Ω ratio across the *Drosophila* clades, but fails to describe the systematic amplitude differences between these clades. In particular, the landscape model drastically overestimates the divergence of close species, $D_{\text{mel-yak}}$ and $D_{\text{mel-sim}}$.
- (b) Ornstein-Uhlenbeck model. In a previous study, Bedford and Hartl [20] analyze the same data set and infer broad stabilizing selection on expression levels, which is consistent

with our results. However, they observe a saturation of gene expression divergence that is at variance with the inference of a linear growth on time scales beyond the divergence time of *D. melanogaster* and *D. simulans* (Fig. 1A and ref. [19]). This can be traced to differences in data analysis. First, Bedford and Hartl [20] use amino acid distances in their phylogeny. These distances are affected by selection [57] and produce relative branch lengths that differ substantially from the phylogeny based on fourfold synonymous sites [42] used in this study (Fig. S1A and Fig. S2). Second, Bedford and Hartl [20] analyze expression divergence for pairs of species, while we group the species into clades (Fig. S1A). These differences lead to a more noisy dependence of the expression divergence data on evolutionary time and make the distinction of conservation and adaptation more difficult (Fig. S2B). Bedford and Hartl [20] fit these data to an Ornstein-Uhlenbeck model of evolution under stabilizing selection [43] (equation 28), which is described below. This model has two independent parameters, which equals the number of fit parameters for our minimal seascape model. Similarly to our landscape model, the best-fit Ornstein-Uhlenbeck model explains the average expression divergence across the *Drosophila* genus (Fig. S2B), but it cannot explain the pattern of expression divergence between close species. The model predicts a quasi-neutral linear growth of the divergence with $D_{mel-yak}/D_{mel-sim} \approx \tau_{mel-yak}/\tau_{mel-sim} \approx 2$ (equation 21), which drastically overestimates the observed ratio $D_{mel-yak}/D_{mel-sim} \approx 1.2$.

Comparison of the Ω test with related methods. Our inference method for selection on quantitative traits can be compared with three well-known selection tests for phenotypic and genomic data:

- (a) Q_{ST}/F_{ST} ratio test for selection on quantitative traits. The summary statistics F_{ST} and Q_{ST} measure the expected fraction of the total genetic variation harboured in a pair of populations that can be attributed to the divergence between these populations; the complementary fraction is attributed to the diversity within populations. F_{ST} refers to neutrally evolving sequence loci [58–60], which can be regarded as a “pseudotrait” with aggregate divergence and diversity. Q_{ST} is the analogous measure for quantitative traits under selection [61]. The expected dependence of these measures on divergence time can be expressed in terms of the Ω ratio (equation 9),

$$F_{ST}(\tau) = \frac{\langle D(\tau) \rangle_0}{\langle D(\tau) \rangle_0 + 2\langle \Delta \rangle_0} = \frac{\Omega_0(\tau)}{\Omega_0(\tau) + 2\pi_0}, \quad (25)$$

$$Q_{ST}(\tau) = \frac{\langle D(\tau) \rangle}{\langle D(\tau) \rangle + 2\langle \Delta \rangle} = \frac{\Omega(\tau)}{\Omega(\tau) + 2\pi_0}, \quad (26)$$

where we use expectation values $\langle \dots \rangle$ in an ensemble of parallel-evolving populations and the subscript 0 refers to neutral evolution. The Q_{ST}/F_{ST} test [29] stipulates that a quantitative trait is evolving at neutrality if $Q_{ST}/F_{ST} = 1$, under stabilizing selection if $Q_{ST}/F_{ST} < 1$, and under directional selection if $Q_{ST}/F_{ST} > 1$. Comparison with the theory of the Ω ratio (equations 17–21) shows that the Q_{ST}/F_{ST} test is insensitive to

selection for divergence times in the quasineutral regime,

$$\frac{Q_{\text{ST}}(\tau)}{F_{\text{ST}}(\tau)} \simeq 1 \quad \text{for } \tau \ll \Omega_{\text{stab}}(c). \quad (27)$$

The data set of this study, which has divergence times $\tau \geq \Omega_{\text{stab}}$ and aggregate values $Q_{\text{ST}}/F_{\text{ST}}$ between 0.6 for the *mel-sim* clade and 0.8 across the entire *Drosophila* genus; these values are obtained using equations (10), (25), and (26). Hence, this test signals broad stabilizing selection but no directional selection. In contrast, the Ω test infers both stabilizing and directional selection from the linear dependence $\Omega(\tau)$ (Fig. 1A and equation 20). This inference shows a conceptually important point: stabilizing and directional selection are not mutually exclusive, but joint features of selection on macro-evolutionary time scales.

- (b) Ornstein-Uhlenbeck model for quantitative trait evolution. This phenomenological model describes a quantitative trait evolving under genetic drift and stabilizing selection [43, 51–53] and has been applied to the evolution of gene expression [20, 44, 45] (a detailed comparison with the results of ref. [20] is given above). The model is defined by a Langevin equation for the population mean trait,

$$\frac{d}{dt}\Gamma(t) = -\lambda(\Gamma - E^*) + \eta_{\Gamma}(t), \quad (28)$$

where $\eta_{\Gamma}(t)$ is the random variable of a delta-correlated Gaussian process with average 0 and variance σ^2 . The model constants λ and σ^2 are usually regarded as independent fit parameters. The Ornstein-Uhlenbeck dynamics of the population mean trait $\Gamma(t)$ around a fixed optimal trait value E^* (equation 28) should not be confused with the Ornstein-Uhlenbeck dynamics of the time-dependent optimum $E^*(t)$ in our seascape model (equation 11). A Langevin equation similar to (28) can be derived from a more general population-genetic model for the evolution of a quantitative trait E a static fitness landscape $f(E) = -c_0(E - E^*)^2$, which has been introduced in ref. [27, 62]. In this model, the population mean trait follows the Ornstein-Uhlenbeck process

$$\frac{d}{dt}\Gamma(t) = -2\langle\Delta\rangle c_0(\Gamma - E^*) - 2\mu(\Gamma - \Gamma_0) + \eta_{\Gamma}(t), \quad (29)$$

where Γ_0 is the genetic mean trait in the long-term limit of neutral evolution and $\eta_{\Gamma}(t)$ is the random variable of a delta-correlated Gaussian process with average 0 and variance $\langle\Delta\rangle/N$. Comparison with equation (28) determines the Ornstein-Uhlenbeck coefficients in terms of the stabilizing strength and the average trait diversity ($\lambda = 2\langle\Delta\rangle c_0$, $\sigma^2 = \langle\Delta\rangle/N$). Equation (29) contains an additional mutational term $(-2\mu)(\Gamma - \Gamma_0)$, which implies that the expectation value $\langle\Gamma\rangle$ differs from the optimum trait value E^* . We note that the diffusion constant $\langle\Delta\rangle/N$ determines the behavior of the Ω ratio (equation 19), of the trait divergence (equation 21), and of the $Q_{\text{ST}}/F_{\text{ST}}$ ratio (equation 27) in the quasineutral regime ($\tau \ll \Omega_{\text{stab}}$). The Ornstein-Uhlenbeck model has been generalized to

account for lineage-specific stabilizing selection in a phylogeny [43–45, 51–53]; however, inferring independent selection parameters for each lineage may lead to overfitting of our data set. Instead, we use the seascape model (11) to infer lineage- and gene-specific changes of the trait optimum $E^*(t)$ using a single additional selection parameter v .

- (c) McDonald-Kreitman test for adaptive sequence evolution [37]. The conceptually closest sequence-based test evaluates the sequence divergence-diversity ratio Ω for a sequence class under putative selection (e.g., non-synonymous mutations in protein-coding sequence) and compares it to the analogous ratio Ω_0 for *bona fide* neutral changes (e.g., synonymous mutations). Positive selection in the query sequence is inferred if $\Omega > \Omega_0$. In this case, the amplitude ratio $\alpha_{\text{seq}} = (\Omega - \Omega_0)/\Omega$ estimates the fraction of non-synonymous substitutions that are adaptive, i.e., driven by positive selection [57]. This is a variant of the McDonald-Kreitman test [37]. It requires data from query sequence and from neutral sequence, but only from a single pair of species with a divergence time beyond the coalescence time. In contrast, the Ω test requires only data from traits under selection, but from three or more species with divergence times beyond the equilibrium relaxation time Ω_{stab} . These differences highlight distinct evolutionary characteristics of quantitative traits. First, such traits have a quasi-neutral regime of macro-evolutionary divergence times (equation 21) that has no direct analogue in sequence evolution [27]. Second, in most cases we do not have a gauge of neutrally evolving traits analogous to synonymous sequence.

Probabilistic inference of selection. Here we describe the extension of our selection inference method to expression data of individual genes. A minimal seascape model is determined by the parameters (c, v) or equivalently by (c, Φ) , where $\Phi = 2cv\tau_{\text{Dros.}}/2N$ denotes the expected cumulative fitness flux over the genus divergence time (equation 16). We derive a posterior probability distribution $Q(c, \Phi | \mathbf{E}^\alpha)$, where $\mathbf{E} = (E_1^\alpha, \dots, E_7^\alpha)$ denotes the expression levels of gene α in the 7 species of our data set. This derivation consists of three steps: we obtain the probability distribution $Q(\mathbf{\Gamma} | c, \Phi)$ of population mean traits $\mathbf{\Gamma}^\alpha = (\Gamma_1^\alpha, \dots, \Gamma_7^\alpha)$ in a given seascape model, we include sampling effects to determine the distribution $Q(\mathbf{E} | \Phi)$, and we use Bayes’ theorem to infer the posterior distribution $Q(c, \Phi | \mathbf{E}^\alpha)$.

The basic building block of evolutionary statistics in the minimal seascape model has been derived previously [26]: the *lineage propagator* $G_\tau(\mathbf{\Gamma}, E^* | \Gamma_a, E_a^*)$ is the probability density of mean and optimal trait values $(\mathbf{\Gamma}, E^*)$, given the values (Γ_a, E_a^*) in an ancestral population at scaled evolutionary distance τ . The lineage propagator is related to the stationary distribution of the seascape dynamics, $Q_{\text{stat}}(\mathbf{\Gamma}, E^*) = \lim_{\tau \rightarrow \infty} G_\tau(\mathbf{\Gamma}, E^* | \Gamma_a, E_a^*)$. Both distributions are Gaussian functions that depend on the seascape model parameters and on the neutral variance (trait scale) E_0^2 ; their detailed analytical form is given in equations (30)–(33) and (A.15)–(A.20) of ref. [26]. The probability distribution of population mean traits across the *Drosophila* genus is the stationary distribution for its last common ancestor multiplied by the lineage propagators for all branches of the phylogeny; this expression is to be integrated

over all unknown expression levels. Specifically, we obtain

$$Q(\mathbf{\Gamma}^\alpha | c, \Phi, E_0^2) = \int Q_{\text{stat}}(\mathbf{\Gamma}_l^\alpha, E_l^*) \prod_{i=1}^{l-1} G_{\tau(i)}(\mathbf{\Gamma}_i^\alpha, E_i^* | \mathbf{\Gamma}_{a(i)}^\alpha, E_{a(i)}^*) d\mathbf{\Gamma}_{k+1}^\alpha \dots d\mathbf{\Gamma}_l^\alpha dE_1^* \dots dE_l, \quad (30)$$

where $i = 1, \dots, k$ labels the extant species and $i = k + 1, \dots, l$ the clade ancestor species (with $l = 2k - 1 = 13$ and the index l referring to the last common ancestor of all species), $a(i)$ denotes the closest ancestor of species i , and $\tau(i)$ is the scaled length of the branch between i and $a(i)$. The deviations of the expression measurements E_i^α from the population mean trait $\mathbf{\Gamma}_i^\alpha$ can be described by a Gaussian sampling error model with variance $(\Delta_i^\alpha/2) + (\delta_i^\alpha/k_i)$, as given by equation (6). We obtain

$$Q(\mathbf{E}^\alpha | c, \Phi, E_0^2) = \frac{1}{Z} \int Q(\mathbf{\Gamma}^\alpha | c, \Phi) \exp \left[-\frac{1}{2} \frac{(E_i^\alpha - \mathbf{\Gamma}_i^\alpha)^2}{(\Delta_i^\alpha/2) + (\delta_i^\alpha/k_i)} \right] d\mathbf{\Gamma}_1^\alpha \dots d\mathbf{\Gamma}_k^\alpha, \quad (31)$$

where Z is a normalization factor. This multi-variate Gaussian integral can be evaluated in a straightforward way by the saddle point method. Here we approximate the noisy diversity values of individual genes by the species averages Δ_i and δ_i . Finally, Bayes' theorem gives the posterior distribution

$$Q(c, \Phi | \mathbf{E}^\alpha) = \frac{\int Q(\mathbf{E}^\alpha | c, \Phi) P_0(c, \Phi) dE_0^2}{\int Q(\mathbf{E}^\alpha | c, \Phi) P_0(c, \Phi) dE_0^2 dc d\Phi}, \quad (32)$$

where $P_0(c, \Phi)$ denotes the prior distribution of seascape parameters. This distribution determines the maximum likelihood posterior values of stabilizing strength, fitness flux, and adaptive fraction of expression divergence,

$$(c^\alpha, \Phi^\alpha) = \arg \max_{c, \Phi} Q(c, \Phi | \mathbf{E}^\alpha), \quad \omega_{\text{ad}}^\alpha(\tau) = \frac{(\tau/\tau_{\text{Dros.}}) \Phi^\alpha}{(\tau/\tau_{\text{Dros.}}) \Phi^\alpha + 1/N}; \quad (33)$$

see equation (24). In equation (32), we use a prior distribution $P_0(c, \Phi) \sim \exp(-ac - b\Phi)$ with Lagrange multipliers a, b that calibrate the average posterior values $\langle c \rangle$ and $\langle \Phi \rangle$ over all genes to our inference from aggregate data (see above). This choice generates a conservative inference of gene-specific seascape parameters that reflects two statistical features of our data. First, gene data \mathbf{E} explained by a seascape model with parameters (c, Φ) and a neutral trait variance E_0^2 (see text above and ref. [27]) have a similar likelihood in a family of models with parameters $(\lambda c, \Phi)$ and neutral trait variance λE_0^2 , where $\lambda > 0$ is a rescaling factor, as long as the stabilizing strength is above some minimum value. In other words, there is a residual freedom in model parameters that leaves the fitness flux Φ invariant. This freedom exists because the gene-specific diversity values Δ_i^α are too noisy to be included in the inference. Our prior distribution favors posterior values c close to the minimum stabilizing strength, which are consistent with the inference from aggregate data. Second, the distribution (30) has an algebraic tail, $Q(\mathbf{E} | c, \Phi) \sim \Phi^{-1}$ for $2N\Phi \gg 1$, which is caused by the diffusive

dynamics of the fitness peak. Our prior distribution suppresses this tail and favors posterior values Φ close to the maximum-likelihood value Φ^* . The validation of this inference scheme by simulation tests is described in section 5.

Statistical significance of the inference. The probabilistic extension of the Ω plays an important role in our global inference: to quantify the statistical significance of our evidence for adaptive evolution under directional selection. Specifically, we evaluate the cumulative log-likelihood score for all genes of our data set as a function of the evolutionary variables c and Φ ,

$$S(c, \Phi) = \sum_{\alpha=1}^g \log Q(c, \Phi | \mathbf{E}^\alpha), \quad (34)$$

where $Q(c, \Phi | \mathbf{E}^\alpha)$ is given by equation (32). This function is shown in Fig. 1C with its maximum shifted to 0. The global maximum-likelihood seascape model has parameters

$$(c^*, \Phi^*) = \arg \max_{c, \Phi} S(c, \Phi) = \left(18.6, \frac{3.6}{2N} \right), \quad (35)$$

compared to $(c_{\text{eq}}^*, 0) = \arg \max_c S(c, \Phi = 0) = (16, 0)$ for the best landscape model and $(c, \Phi) = (0, 0)$ for neutral evolution. We obtain log-likelihood differences $S(c^*, \Phi^*) - S(c_{\text{eq}}^*, 0) = 8396$ and $S(c^*, \Phi^*) - S(0, 0) = 12464$. By a log-likelihood test, these differences translate into the P values quoted in the main text. Maximum-likelihood values analogous to equation (35) can also be defined for classes of genes (Table 1).

3. Analysis of alternative evolutionary scenarios

Lineage-specific demography. Demographic effects, such as population bottlenecks, affect the patterns of sequence variation in *Drosophila* [30, 31, 33, 34, 63]. Here we examine the effects of strong, long-term demographic heterogeneities on the divergence and diversity of expression levels. Specifically, we consider changes in effective population size to a value $N_i = \lambda N$ in a given *Drosophila* lineage i , which is defined by the terminal branch of species i in the phylogeny and extends over a scaled evolutionary time τ_i (Fig. S1A). A depletion of effective population size leads to a global relaxation of stabilizing selection on gene expression, given by a reduced stabilizing strength λc in the fitness seascape (equation 11). For each clade \mathcal{C} with $i \in \mathcal{C}$, we define the polarized divergence-diversity ratio,

$$\Omega_{\mathcal{C}, i} = \frac{1}{|\mathcal{C} \setminus \mathcal{C}_1|} \sum_{j \in \mathcal{C} \setminus \mathcal{C}_1} \Omega_{ij}, \quad (36)$$

where $(\mathcal{C}_1, \mathcal{C} \setminus \mathcal{C}_1)$ is the partitioning of clade \mathcal{C} defined by its root node and we assume $i \in \mathcal{C}_1$. The pairwise ratios Ω_{ij} are given by equation (9). Similarly, we define the polarized

divergence time,

$$\tau_{\mathcal{C},i} = \frac{1}{|\mathcal{C} \setminus \mathcal{C}_1|} \sum_{j \in \mathcal{C} \setminus \mathcal{C}_1} \tau_{ij}. \quad (37)$$

In Fig. S3A,B, we plot polarized data $(\tau_{\mathcal{C},i}, \Omega_{\mathcal{C},i})$ together with background data $(\tau_{\mathcal{C}}, \Omega_{\mathcal{C}})$ from partial clades excluding species i . Under a change of population size in lineage i with $\tau_i \gtrsim \Omega_{\text{stab}}(\lambda c)$, the polarized data should follow a pattern with reduced ($\lambda < 1$) or increased ($\lambda > 1$) long-term constraint,

$$\begin{aligned} \Omega(\tau, \tau_i) &= \Omega_{\text{eq}}(\tau, \tau_i) + \Omega_{\text{ad}}(\tau, \tau_i) \\ &= \begin{cases} \tau & \text{for } \tau \ll \Omega_{\text{stab}}(\lambda c) \\ \frac{1}{2}(\Omega_{\text{stab}}(\lambda c) + \Omega_{\text{stab}}(c)) + \frac{1}{2}v\tau + \mathcal{F}(\lambda, c) & \text{for } \tau \gg \tau_i + \Omega_{\text{stab}}(c), \end{cases} \end{aligned} \quad (38)$$

where the shift $\mathcal{F}(\lambda, c)$ is generated by the demographic inhomogeneity on intermediate time scales; this pattern is shown in Fig. S3A,B for $\lambda = 1/2$ and $\lambda = 3$. A similar calculation shows that short-term population bottlenecks have a negligible effect on the Ω statistics. We observe no deviation between polarized and background Ω data, indicating the absence of strong demographic effects shaping the evolution of expression levels. Equation (38) also shows that demographic effects do not confound the Ω test for adaptive evolution under directional selection. For time-independent optimal trait value ($v = 0$), global relaxation of stabilizing selection increases the divergence-diversity ratio, as noted in previous studies [5, 16, 17]; however, it does not generate the linear increase $\Omega_{\text{ad}}(\tau) \simeq v\tau/2$ characteristic of fitness peak displacements (Fig. S3).

Gene-specific relaxation of stabilizing selection. We can also test for lineage- and gene-specific relaxation of stabilizing selection on gene expression, which arises, for example, from a partial loss of gene function. We model the loss dynamics by a stochastic process: with a small rate γ , individual genes switch the stabilizing strength of their fitness seascape to a reduced value λc (with $\lambda < 1$). This dynamics increases cross-species divergence and generates a nonlinear time-dependent $\Omega(\tau)$, not observed in the data (Fig. 1A). To discriminate between relaxed stabilizing selection and directional selection, we also use the distributions of clade-specific expression differences $\Delta E_{\mathcal{C}}^{\alpha}$, which are defined as averages over pairwise differences $\Delta E_{ij}^{\alpha} = E_i^{\alpha} - E_j^{\alpha}$ in analogy to equation (10). The observed distributions are of approximately Gaussian form,

$$P_{\mathcal{C}}(\Delta E) = \frac{1}{\sqrt{2\pi D_{\mathcal{C}}}} \exp \left[-\frac{(\Delta E)^2}{2D_{\mathcal{C}}} \right], \quad (39)$$

as shown by the collapse plot of Fig. S4A. This is in accordance with the minimal seascape model, which predicts a Gaussian distribution $P_{\tau}(\Delta E)$ with variance $\langle D(\tau) \rangle$. In contrast, stochastic relaxation of stabilizing selection generates broad non-Gaussian tails increasing with divergence time τ that are not observed in the data (Fig. S4C). We conclude that relaxed

stabilized selection alone cannot explain the observed statistics of *Drosophila* gene expression levels. This does not exclude that relaxation of selection affects some genes in our data set and more broadly genes with complete loss of function, which are suppressed in the set of conserved orthologs.

Punctuated directional selection. The Ornstein-Uhlenbeck dynamics of fitness peaks in the minimal seascape model (equation 28) describes the accumulation of small but continual changes of optimal expression levels. Larger peak shifts can be caused by discrete ecological events, including major migrations and speciations, and by gene-specific factors such as neo-functionalization [64]. Here we model such events by a *punctuated* fitness seascape [26]: with a small rate $v\mu/(2r^2)$, individual genes are subject to fitness peak shifts by an amount of order E_0 . This stochastic model differs from previous models of lineage-specific selection [20, 24, 43–45, 51, 52], where fitness peak shifts are constrained to known branch points of the phylogeny. Evolution in a punctuated fitness seascape generates divergence-diversity ratio $\Omega(\tau)$ of the form (equation 20); adaptation is signalled by the same term $\Omega_{\text{ad}}(\tau) \simeq v\tau/2$ as in a minimal seascape of the same driving rate v . To discriminate between the two models, we use again the distributions $P_C(\Delta E)$ of clade-specific expression differences. In a punctuated seascape, these distributions have broad non-Gaussian tails increasing with divergence time τ_C that are not observed in the data (Fig. S4D). We conclude that large peak shifts are a subleading factor of expression changes in our data set.

Other modes of adaptation. Further evolutionary modes affecting gene expression include:

- (a) Time-dependent stabilizing selection [26]. This type of selection can be modeled by a fitness seascape of the form (11) with time-dependent stabilizing strength $c(t)$, given by a generalized Ornstein-Uhlenbeck process with constraint $c(t) > c_{\text{min}}$. The recurrent tightening of expression constraint driven by increases of $c(t)$ is a mode of adaptation that is independent of fitness peak changes. The Ω test does not trace this mode: as long as the expression optimum E^* is time-independent, the function $\Omega(\tau)$ reaches an asymptotic value $< \Omega_{\text{stab}}(c_{\text{min}})$. This pattern is similar to evolutionary equilibrium in a single-peak fitness landscape and does not contain the term $\Omega_{\text{ad}}(\tau) \simeq v\tau/2$ characteristic of fitness peak displacements.
- (b) Adaptive gene turnover, including sub- and neo-functionalization after gene duplication [65, 66], regulatory sequence duplication [67], and de novo formation of genes [68]. This mode is suppressed in our data set of conserved orthologous genes, but it is likely to be more prevalent in the complementary set of *Drosophila* genes.

A detailed investigation of these evolutionary modes is beyond the scope of this study. Importantly, however, they do not confound the inference of adaptation under directional selection reported here.

4. Analysis of specific gene classes

Codon usage. The effective number of codons, n , measures the redundancy of the genetic code within a given gene [46]. This number takes values between 20 (each amino acid is determined by a specific codon) and 61 (all sense codons are used). Genes with specific codon usage (small n) tend to have higher expression than genes with broad codon usage [35, 36]. Here we compute the species-averaged effective number of codons, $n^\alpha = \frac{1}{7} \sum_i n_i^\alpha$ for all genes in our data set. We find a consistent dependence of expression adaptation on codon usage:

- (a) Aggregate analysis by the Ω test signals strongly reduced adaptation for genes with specific codon usage ($n < 42$) and an enhanced adaptation for genes with broad codon usage ($n > 50$), compared to the average over all genes (Fig. 2A and Table 1).
- (b) The Ω test also signals strongly reduced adaptation for genes with high average expression level, $\bar{E}^\alpha = \frac{1}{7} \sum_i E_i^\alpha > 0.9$ (Table 1). Additionally, we compare the fitness flux of a gene to its codon adaptation index (CAI), which measures the similarity between the codon usage in a specific gene and the codon preference of highly expressed genes [69]. Consistently, we find a reduced amount of fitness flux in genes with high codon adaptation index ($\text{CAI} \gtrsim 0.65$); these genes are likely to be highly expressed.
- (c) At the level of individual genes, there is a clear correlation between fitness flux Φ^α and effective codon number n^α (Fig. 2B).

Inference of adaptive sequence evolution. For the genes in our data set, we estimate the fractions of synonymous and non-synonymous polymorphic nucleotides (P_s and P_n) from the database of *Drosophila melanogaster* Genetic Reference Panel (DGRP) [41]. The corresponding nucleotide divergence measures (D_s and D_n) are obtained from sequence alignments between the *D. melanogaster* and *D. simulans* reference genomes [42]. The McDonald-Kreitman test [37, 57] signals adaptive evolution of amino acids if $\alpha_{\text{seq}} = (D_n P_s / D_s P_n) - 1 > 0$. Fig. S5 shows the distribution of α_{seq} values for classes of genes with different amount of expression adaptation, measured by the fitness flux Φ (equation 33). We find no correlation between these statistics. In each class, about 30% of the genes have $\alpha_{\text{seq}} > 0$. This result does not contradict the correlation of gene expression divergence and amino acid divergence reported in ref. [19], because an enhanced amino acid substitution rate measured by a D_n/D_s test [70] may be caused by adaptive changes or by relaxation of negative selection.

Analysis of functional gene classes. We use The Ontologizer [71] for statistical analysis of functional enrichment in our dataset. From a base set of all 6332 genes in our database, we identify enriched functional categories in the query sets of adaptively regulated genes ($2N\Phi^\alpha > 4$) and genes with sex-specific adaption of expression ($2N\Phi_{\text{mf}}^\alpha > 4.5$, see below). We use the calculation method Parent-Child-Union with Bonferroni correction and resampling steps of 1000. The enriched functional categories in these gene sets are reported in Tables S1 and S2 with a significance threshold $P < 0.1$ (multiple hypothesis test). We list

three main categories: biological processes, cellular components, and molecular functions. Each functional category is assigned to a functional cluster (in bold letters) that is inferred by REVIGO [72], using the semantic similarity measure SimRel with threshold 0.5. This clustering facilitates the interpretation of functional gene classes associated with adaptation of gene expression.

Sex-specific evolution and sex bias of expression. To quantify differences of gene expression between male and female individuals, we define the sex specificity trait of a given gene as the difference between its expression levels in males and in females [19],

$$E_{\text{mf},i}^{\alpha} = E_{m,i}^{\alpha} - E_{f,i}^{\alpha}. \quad (40)$$

We analyze these traits by the same methods as the sex-averaged expression levels E_i^{α} defined by equation (5). Specifically, we define the aggregate divergence-diversity ratio $\Omega_{\text{mf},\mathcal{C}}$ and the fitness flux $2N\Phi_{\text{mf}}$ in analogy to equations (10) and (15), and we infer gene-specific maximum-likelihood values $2N\Phi_{\text{mf}}^{\alpha}$ in analogy to equation (11). We define two conceptually distinct measures of male-female differentiation:

- (a) Sex-specific adaptation. In accordance with ref. [19], we find that most genes of our data set have well-conserved and often small sex specificity; these genes evolve their expression levels coherently between males and females. We use the rescaled fitness flux $2N\Phi_{\text{mf}}$ to delineate coherent evolution of expression levels (i.e., conservation of the specificity trait) from sex-specific adaptation (i.e., adaptive changes of the male-female expression difference), as illustrated in Fig. 2C. A set of 1155 sex-specific adaptive genes is identified by the condition $2N\Phi_{\text{mf}}^{\alpha} > 4.5$ (Table S2); we use a more stringent threshold than for Φ^{α} because the sex-specificity trait statistics has larger statistical errors.
- (b) Sex bias. We identify genes with male- and female-biased expression in *Drosophila* using the results of Assis et al. [39], which are based on a number of statistical tests in the whole body and in gonads of males and females in *D. melanogaster* and *D. pseudoobscura*. A gene is classified as expression sex-biased if flagged by at least three of these tests, which produces a list of 450 male-biased and 499 female-biased genes. A related measure of bias within our data set is the species-averaged specificity trait, $\bar{E}_{\text{mf}}^{\alpha} = \frac{1}{7} \sum_i E_{\text{mf},i}^{\alpha}$.

Our analysis establishes a relation between these two measures in our data set: strong sex-specific adaptation of expression occurs in male-biased, but not in female-biased genes. First, the aggregate ratio Ω_{mf} in male-biased genes show evidence for adaptive evolution with a linear adaptive component $v\tau/2$. Unbiased and female-biased genes have only a small average divergence in their sex-specificity trait that is of the order of the expression diversity (i.e., they within the error range of the observed expression levels), providing no evidence for adaptation (Fig. 2D). Second, the fitness flux $\Phi_{\text{mf}}^{\alpha}$ is strongly enhanced for genes with large $\bar{E}_{\text{mf}}^{\alpha}$ (Fig. 2E). Accordingly, 32% of male-biased genes are classified as sex-specific adaptive. Functional categories associated with sex-specific adaptation of expression are reported in Table S2.

5. Simulation tests

In-silico evolution of quantitative traits. We use a Fisher-Wright process for the evolution for the evolution of populations along the *Drosophila* phylogeny of Fig. S1A. A population consists of N individuals with genomes $\mathbf{a}^{(1)}, \dots, \mathbf{a}^{(N)}$. A genotype is an ℓ -letter sequence $\mathbf{a} = (a_1, \dots, a_\ell)$ with alleles $a_k = 0, 1$ ($k = 1, \dots, \ell$). It defines an expression level $E(\mathbf{a}) = \sum_{k=1}^{\ell} \mathcal{E}_k a_k$ with neutral variance $E_0^2 = \frac{1}{4} \sum_{k=1}^{\ell} \mathcal{E}_k^2$. We use uniform single-locus effects \mathcal{E}_i ; our results are insensitive to the form of the effect distribution [26]. In each generation, the sequences undergo point mutations with a probability $\mu\tau_0$ per generation, where τ_0 is the generation time. The sequences of next generation are then obtained by multinomial sampling with a probability proportional to $[1 + \tau_0 f(E(\mathbf{a}), t)]$, where the fitness function $f(E, t)$ is given by equation (11). Simulations are performed with $N = 100$, $\pi_0 = 0.1$ for traits with $\ell = 100$, uniform effects $E_i = 1$, and average fitness optimum $\mathcal{E} = 70$. We use three different types of selection (for details, see ref. [26]):

- (a) Minimal fitness seascape. Before each reproduction step, a new optimal trait value $E^*(t + \tau_0)$ is drawn from a Gaussian distribution with mean $E^*(t)(1 - \mu\tau_0 v/(2r^2)) + \mathcal{E} \mu\tau_0 v/(2r^2)$ and variance $\mu\tau_0 v E_0^2$.
- (b) Fitness landscape. The optimal trait value E^* is time-independent (Fig. S3A,B). In the model of gene-specific relaxation of selection (see section 3), the stabilizing strength of individual genes switches to a smaller value, $c \rightarrow 0.01c$, with a small rate γ (Fig. S4C).
- (c) Punctuated fitness seascape (see section 3). Before each reproduction step, a new, uncorrelated optimal trait value is drawn with probability $\mu\tau_0 v/(2r^2)$ from a Gaussian distribution with variance $r^2 E_0^2$, where r^2 is a constant of order 1 (Fig. S4D).

Validation of the probabilistic inference scheme. To test the performance of our inference scheme, we generate expression values $\mathbf{E}^\alpha = (E_1^\alpha, \dots, E_7^\alpha)$ for individual genes with trait scales $E_{0,\alpha}^2$ by Fisher-Wright simulations along the *Drosophila* phylogeny of Fig. S1A. We use minimal fitness seascapes of the form (11) with input parameters $(c_{\text{in}}, v_{\text{in}})$ and a sequence diversity $\pi_0 = 4\mu N = 0.05$. We then infer maximum-likelihood posterior values $(c^\alpha, 2N\Phi^\alpha)$ by the probabilistic method described in section 2 (equation 33). In Fig. S7A, we plot the distribution of inferred fitness flux values $2N\Phi^\alpha$ against the input expectation value $2N\Phi_{\text{in}} = 2c_{\text{in}}v_{\text{in}}\tau_{\text{Dros.}}$ (equation 16). The underlying simulations use a range of trait scales $E_{0,\alpha}^2 = 0.25 - 4.0$ appropriate for log expression levels; the inference of Φ^α does not require knowledge of this scale (see section 2). Fig. S7B shows the corresponding distribution of inferred values c^α as a function of the input stabilizing strength c_{in} . These simulations use a uniform trait scale $E_{0,\alpha}^2 = 1$ (inferring the actual scales requires sufficiently reliable gene-specific expression diversity data). The posterior values (Φ^α, c^α) are seen to provide reasonable, on average conservative estimates of the input model parameters $(c_{\text{in}}, \Phi_{\text{in}})$. In particular, the inference of a significant fitness flux ($\Phi > 1/2N$) is incompatible with evolution under static stabilizing selection ($v = 0, c > 0$) or near neutrality ($c \simeq 1$), independently of the underlying model for the adaptive evolution of a molecular trait.

Robustness of the inference to trait epistasis. The analytical theory underlying our inference method [26, 27] covers molecular quantitative traits with a linear genotype-phenotype map, $E(\mathbf{a}) = \sum_{k=1}^{\ell} \mathcal{E}_k a_k$ (see above). Here we extend this method to nonlinear traits of the form $E(\mathbf{a}) = \sum_{k=1}^{\ell} \mathcal{E}_k a_k + \sum_{k < k'} \mathcal{E}_{kk'} a_k a_{k'}$; such nonlinearities are commonly referred to as trait epistasis. The strength of epistasis can be defined as the ratio of nonlinear and linear neutral trait variance, $\varepsilon^2 = (\sum_{k < k'} \mathcal{E}_{kk'}^2) / (\sum_k \mathcal{E}_k^2)$.

Trait epistasis introduces only minor changes to the quantitative genetics theory of refs. [26, 27]. In particular, the Ω ratio retains its normalization in the neutral long-term limit and the quasi-neutral growth of the trait divergence is still given by equation (21), where Δ is now the total genetic diversity of the trait. To specifically test our inference method, we perform Fisher-Wright simulations as described above over a wide range of the parameter ε^2 ; individual epistatic effects $\mathcal{E}_{kk'}$ are drawn from a Gaussian distribution with mean 0. In an ensemble of 6000 independently evolving traits, we record both the actual average fitness flux (equation 15) and the inferred fitness flux determined from the aggregate Ω ratio (equation 24). Both quantities show no systematic dependence on ε^2 (Fig. S7C), suggesting that our inference of adaptive evolution is not confounded by trait epistasis.

Acknowledgments. We acknowledge discussions with P. Andolfatto, N. Barton, M. Luksza, V. Mustonen, P. Shah, and S. Schiffels. This work has been partially supported by James S. McDonnell Foundation 21st century science initiative-postdoctoral program in complexity science / complex systems (AN) and and by Deutsche Forschungsgemeinschaft grant SFB 680. This work has also been supported by the Gordon and Betty Moore Foundation grant no. 2919 and National Institutes of Health grant no. R25 GM067110 to the Kavli Institute of Theoretical Physics (Santa Barbara), where part of this work was performed.

References

1. P. Andolfatto. Adaptive evolution of non-coding DNA in *Drosophila*. *Nature*, 437(7062):1149–1152, 2005.
2. V. Mustonen and M. Lässig. Adaptations to fluctuating selection in *Drosophila*. *Proceedings of the National Academy of Sciences*, 104(7):2277–2282, 2007.
3. G. Sella, D. A. Petrov, M. Przeworski, and P. Andolfatto. Pervasive natural selection in the *Drosophila* genome? *PLoS Genetics*, 5(6):e1000495, 2009.
4. M. King and A. Wilson. Evolution at two levels in humans and chimpanzees. *Science*, 188(4184):107–116, 1975.
5. H. B. Fraser. Genome-wide approaches to the study of adaptive gene expression evolution. *Bioessays*, 33(6):469–477, 2011.
6. I. G. Romero, I. Ruvinsky, and Y. Gilad. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, 13(7):505–516, 2012.

7. A. A. Pai, J. K. Pritchard, and Y. Gilad. The genetic and mechanistic basis for variation in gene regulation. *PLoS Genetics*, 11(1):e1004857, 2015.
8. H. B. Fraser, A. M. Moses, and E. E. Schadt. Evidence for widespread adaptive evolution of gene expression in budding yeast. *Proceedings of the National Academy of Sciences*, 107(7):2977–2982, 2010.
9. A. Genissel, L. M. McIntyre, M. L. Wayne, and S. V. Nuzhdin. Cis and trans regulatory effects contribute to natural variation in transcriptome of *Drosophila melanogaster*. *Molecular Biology and Evolution*, 25(1):101–110, 2007.
10. P. J. Wittkopp, B. K. Haerum, and A. G. Clark. Regulatory changes underlying expression differences within and between *Drosophila* species. *Nature Genetics*, 40(3):346–350, 2008.
11. J. D. Coolon, C. J. McManus, K. R. Stevenson, B. R. Graveley, and P. J. Wittkopp. Tempo and mode of regulatory evolution in *Drosophila*. *Genome Research*, 24(5):797–808, 2014.
12. C. G. Artieri and H. B. Fraser. Evolution at two levels of gene expression in yeast. *Genome Research*, 24(3):411–21, 2014.
13. H. E. Hoekstra and J. A. Coyne. The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, 61(5):995–1016, 2007.
14. S. A. Rifkin, J. Kim, and K. P. White. Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nature Genetics*, 33(2):138–144, 2003.
15. B. Lemos, C. D. Meiklejohn, M. Cáceres, and D. L. Hartl. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution*, 59(1):126–137, 2005.
16. P. Khaitovich. Parallel patterns of evolution in the genomes and transcriptomes of humans and chimpanzees. *Science*, 309(5742):1850–1854, 2005.
17. Y. Gilad, A. Oshlack, G. K. Smyth, T. P. Speed, and K. P. White. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature*, 440(7081):242–245, 2006.
18. A. Whitehead and D. L. Crawford. Neutral and adaptive variation in gene expression. *Proceedings of the National Academy of Sciences*, 103(14):5425–5430, 2006.
19. Y. Zhang, D. Sturgill, M. Parisi, S. Kumar, and B. Oliver. Constraint and turnover in sex-biased gene expression in the genus *Drosophila*. *Nature*, 450(7167):233–237, 2007.
20. T. Bedford and D. L. Hartl. Optimization of gene expression by natural selection. *Proceedings of the National Academy of Sciences*, 106(4):1133–1138, 2009.
21. H. B. Fraser, T. Babak, J. Tsang, Y. Zhou, B. Zhang, M. Mehrabian, and E. E. Schadt. Systematic detection of polygenic cis-regulatory evolution. *PLoS Genetics*, 7(3):e1002023, 2011.
22. M. Lynch and W. G. Hill. Phenotypic evolution by neutral mutation. *Evolution*, 1986.

23. R. Blekhman, A. Oshlack, A. E. Chabot, G. K. Smyth, and Y. Gilad. Gene regulation in primates evolves under tissue-specific selection pressures. *PLoS Genetics*, 4(11):e1000271, 2008.
24. T. Brawand, M. Soumillon, A. Necsulea, P. Julien, G. Csárdi, et al. The evolution of gene expression levels in mammalian organs. *Nature*, 478(7369):343–8, 2011.
25. H. Fraser. Gene expression drives local adaptation in humans. *Genome Research*, 23(7):1089–96, 2013.
26. T. Held, A. Nourmohammad, and M. Lässig. Adaptive evolution of molecular phenotypes. *Journal of Statistical Mechanics: Theory and Experiment*: P09029, 2014.
27. A. Nourmohammad, S. Schiffels, and M. Lässig. Evolution of molecular phenotypes under stabilizing selection. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01012, Jan. 2013.
28. A. Nourmohammad, T. Held, and M. Lässig. Universality and predictability in molecular quantitative genetics. *Current Opinion in Genetics and Development*, 23(6):684–693, 2013.
29. T. Leinonen, R. J. S. McCairns, R. B. O’Hara, and J. Merilä. Q(ST)-F(ST) comparisons: evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics*, 14(3):179–190, 2013.
30. D. Lachaise, M. L. Cariou, J. R. David, and F. Lemeunier. Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evolutionary Biology*, 22:159–225, 1988.
31. S. Glinka, L. Ometto, S. Mousset, W. Stephan, and D. De Lorenzo. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. *Genetics*, 165(3):1269–1278, 2003.
32. P. R. Haddrill, K. R. Thornton, B. Charlesworth, and P. Andolfatto. Multilocus patterns of nucleotide variability and the demographic and selection history of *Drosophila melanogaster* populations. *Genome Research*, 15(6):790–9:2005.
33. W. Stephan and H. Li. The recent demographic and adaptive history of *Drosophila melanogaster*. *Heredity*, 98(2):65–68, 2007.
34. K. R. Thornton, J. D. Jensen, C. Becquet, and P. Andolfatto. Progress and prospects in mapping recent selection in the genome. *Heredity*, 98(6):340–348, 2007.
35. T. Ikemura. Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, 2(1):13–34, 1985.
36. D. C. Shields, P. M. Sharp, D. G. Higgins, and F. Wright. Silent sites in *Drosophila* genes are not neutral - Evidence of selection among synonymous codons. *Molecular Biology and Evolution*, 5(6):704–716, 1988.
37. J. H. McDonald and M. Kreitman. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature*, 351(6328):652–654, 1991.

38. A. M. Hancock, D. B. Witonsky, G. Alkorta-Aranburu, C. M. Beall, A. Gebremedhin, et al. Adaptations to climate-mediated selective pressures in humans. *PLoS Genetics*, 7(4):e1001375, 2012.
39. R. Assis, Q. Zhou, and D. Bachtrog. Sex-biased transcriptome evolution in *Drosophila*. *Genome Biology and Evolution*, 4(11):1189–1200, 2012.
40. S. R. Sunyaev, F. P. Roth. Systems biology and the analysis of genetic variation. *Current Opinion in Genetics and Development*, 23(6):599–601, 2013.
41. T. F. C. Mackay, S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, et al. The *Drosophila melanogaster* Genetic Reference Panel. *Nature*, 482(7384):173–178, 2012.
42. *Drosophila* 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*, 450(7167):203–218, Nov. 2007.
43. T. F. Hansen. Stabilizing selection and the comparative analysis of adaptation. *Evolution*, 51(5):1341–1351, 1997.
44. A. T. Kalinka, K. M. Varga, D. T. Gerrard, S. Preibisch, D. L. Corcoran, et al. Gene expression divergence recapitulates the developmental hourglass model. *Nature*, 468(7325):811–814, 2010.
45. R. V. Rohlf, P. Harrigan, and R. Nielsen. Modeling gene expression evolution with an extended Ornstein-Uhlenbeck process accounting for within-species variation. *Molecular Biology and Evolution*, 31(1):201–211, 2014.
46. F. Wright. The ‘effective number of codons’ used in a gene. *Gene*, 87(1):23–29, 1990.
47. Y. Gilad, A. Oshlack, and S. A. Rifkin. Natural selection on gene expression. *Trends in Genetics*, 22(8):456–461, 2006.
48. A. Tsankov, D. A. Thompson, A. Socha, A. Regev, and O. J. Rando. The role of nucleosome positioning in the evolution of gene regulation. *PLoS Biology*, 8(7):e1000414, 2010.
49. G. H. Perry, P. Melsted, J. C. Marioni, Y. Wang, R. Bainer, et al. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Research*, 22(4):602–610, 2012.
50. J. Quackenbush. Microarray data normalization and transformation. *Nature Genetics*, 32 Suppl:496–501, 2002.
51. M. A. Butler and A. A. King. Phylogenetic comparative analysis: A modeling approach for adaptive evolution. *American Naturalist*, 164(6):683–695, 2004.
52. T. F. Hansen, J. Pienaar, and S. H. Orzack. A comparative method for studying adaptation to a randomly evolving environment. *Evolution*, 62(8):1965–1977, 2008.
53. J. M. Beaulieu, D. C. Jhvueng, C. Boettiger, and B. C. O’Meara. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution*, 66(8):2369–2383, 2012.

54. V. Mustonen and M. Lässig. Fitness flux and ubiquity of adaptive evolution. *Proceedings of the National Academy of Sciences*, 107(9):4248–4253, 2010.
55. R. Chakraborty and M. Nei. Genetic differentiation of quantitative characters between populations or species .1. mutation and random genetic drift. *Genetical Research*, 39(3):303–314, 1982.
56. M. Lynch and B. Walsh. *Genetics and analysis of quantitative traits*. Sinauer Associates Inc, 1998.
57. N. G. C. Smith and A. Eyre-Walker. Adaptive protein evolution in *Drosophila*. *Nature*, 415(6875):1022–1024, 2002.
58. S. Wright. Isolation by distance. *Genetics*, 1943.
59. S. Wright. Genetical structure of populations. *Nature*, 166(4215):247–249, 1950.
60. R. Lande. Neutral theory of quantitative genetic variance in an island model with local extinction and colonization. *Evolution*, 46(2):381, 1992.
61. K. Spitze. Population structure in *Daphnia obtusa*: quantitative genetic and allozymic variation. *Genetics*, 135(2):367–374, 1993.
62. H. P. de Vladar and N. H. Barton. The statistical mechanics of a polygenic character under stabilizing selection, mutation and drift. *Journal of the Royal Society Interface*, 8(58):720–739, 2011.
63. C. F. Aquadro, V. Bauer DuMont, and F. A. Reed. Genome-wide variation in the human and fruitfly: a comparison. *Current Opinion in Genetics and Development*, 11(6):627–634, 2011.
64. M. Lynch and A. Force. The probability of duplicate gene preservation by subfunctionalization. *Genetics*, 154(1):459–473, 2000.
65. M. Lynch, M. O’Hely, B. Walsh, and A. Force. The probability of fixation of a newly arisen gene duplicate. *Genetics*, 159(4):1789–1804, 2001.
66. M. Lynch and V. Katju. The altered evolutionary trajectories of gene duplicates. *Trends in Genetics*, 20(11):544–549, 2004.
67. A. Nourmohammad and M. Lässig. Formation of regulatory modules by local sequence duplication. *PLoS Computational Biology*, 7(10):e1002167, 2011.
68. D. Tautz and T. Domazet-Lošo. The evolutionary origin of orphan genes. *Nature Reviews Genetics*, 12(10): 692-702, 2011.
69. P. M. Sharp and W. H. Li. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, 15(3):1281–1295, 1987.
70. W. H. Li. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. *Journal of Molecular Evolution*, 36(1):96–9, 1993.

71. S. Bauer, S. Grossmann, M. Vingron, and P. N. Robinson. Ontologizer 2.0—a multifunctional tool for term enrichment analysis and data exploration. *Bioinformatics (Oxford, England)*, 24(14):1650–1651, 2008.
72. F. Supek, M. Bošnjak, N. Škunca, and T. Šmuc. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One*, 6(7):e21800, 2011.

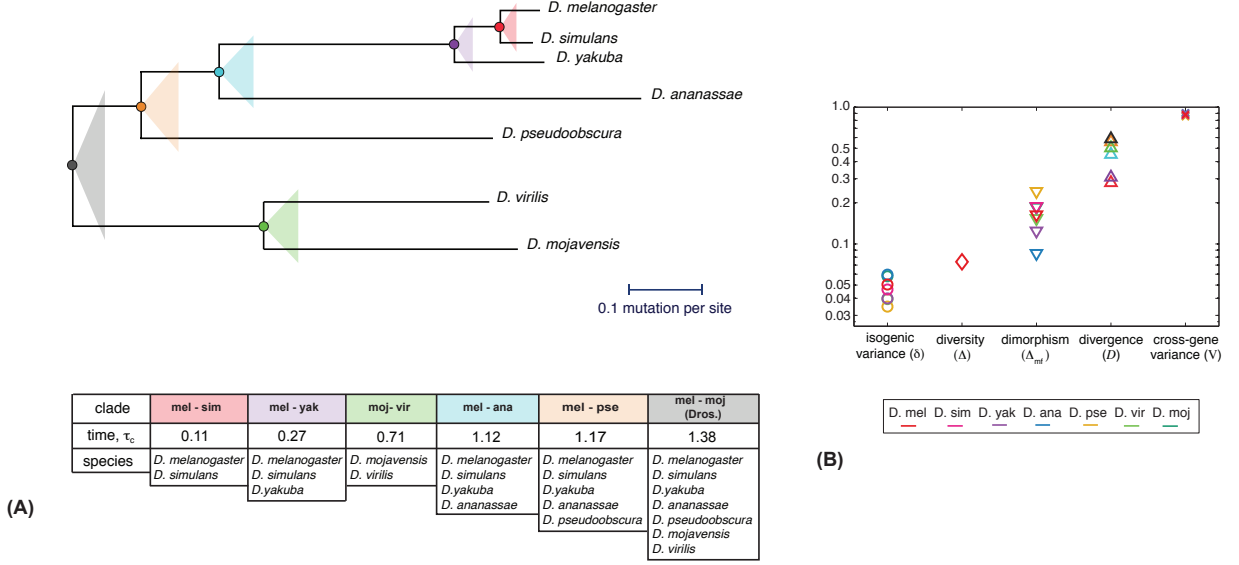


Figure S1: Species phylogeny and levels of gene expression variation. (A) Phylogeny of the *Drosophila* genus, as reconstructed in ref. [42] from synonymous sequence divergence. Six clades are marked by colored triangles, their ancestral nodes by colored circles. The table specifies the species contained each of the clades and the clade divergence time τ_c (equation 1). (B) Gene-averaged isogenic variance $\langle \delta \rangle$ (\circ , equation 4), expression diversity $\langle \Delta \rangle$ (∇ , equation 6), male-female expression dimorphism $\langle \Delta_{mf} \rangle$ (\diamond , equation 7), clade divergence $\langle D \rangle$ (\triangle , equation 8), and cross-gene variance of expression $V = \langle \Gamma_i^2 \rangle \approx 1$ (\times). We find a clear ranking $\langle \delta_i \rangle < \langle \Delta \rangle \lesssim \langle \Delta_{mf} \rangle < \langle D_{ij} \rangle < V_i$. The color code for single-species data is shown in the legend, colors for clades are as in (A).

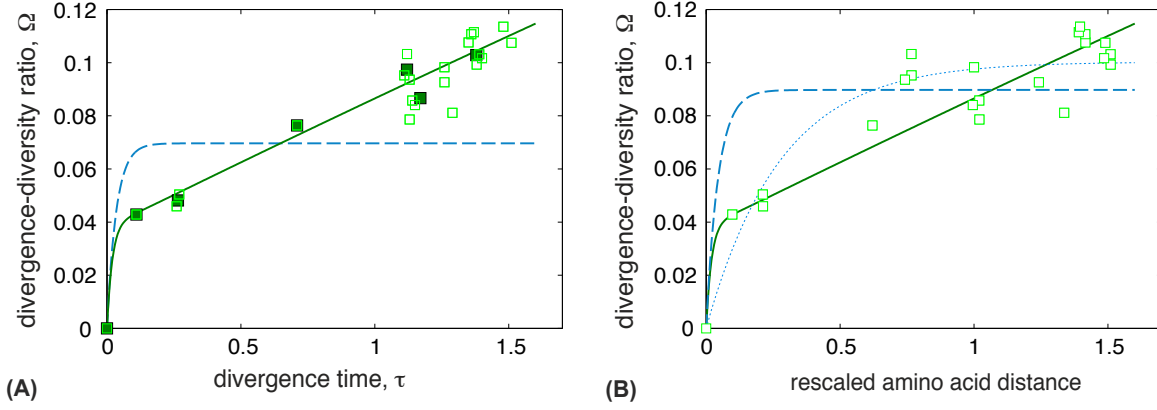


Figure S2: Fitness landscape and Ornstein-Uhlenbeck models as control. (A) The aggregate divergence- diversity ratio for clades, Ω_C (filled squares), is compared to the ratio for individual pairs of species, Ω_{ij} (empty squares). Both quantities are plotted against the divergence time estimated from four-fold synonymous sites [42] (Fig. S1A). Clade-based statistics is seen to substantially reduce the noise of the expression divergence data. The seascape model (green line as in Fig. 1A; $c^* = 18.6$, $v^* = 0.07$) provides a significantly better fit to these data than the landscape model (blue line; $c_{eq} = 13$); see Fig. 1C and section 2 of Materials and Methods for a statistical model comparison. (B) The same species-pair ratio Ω_{ij} (empty squares) is plotted against the amino-acid sequence distance used in ref. [20], uniformly rescaled to give the same scaled genus divergence time $\tau_{Dros.} = 1.4$ as in (A). Compared to synonymous divergence, amino-sequence divergence is seen to produce an inhomogeneous molecular clock that adds to the noise of the evolutionary analysis. For these data, the seascape model (green line as in (A)), the landscape model (dashed blue line; $c = 6.4$), and the Ornstein-Uhlenbeck model [43] (blue line; $\lambda = 1.8 \mu^{-1}$, $\sigma^2 = 0.13 \langle \Delta \rangle / N$; see equation 28 and ref. [20]) produce fits of overall comparable quality. However, both equilibrium models cannot explain the evolution of expression in the youngest clades: the landscape model overestimates the divergence $D_{mel-yak}$ and $D_{mel-sim}$, the Ornstein-Uhlenbeck model overestimates the relative divergence $D_{mel-yak} / D_{mel-sim}$. See section 2 of Materials and Methods.

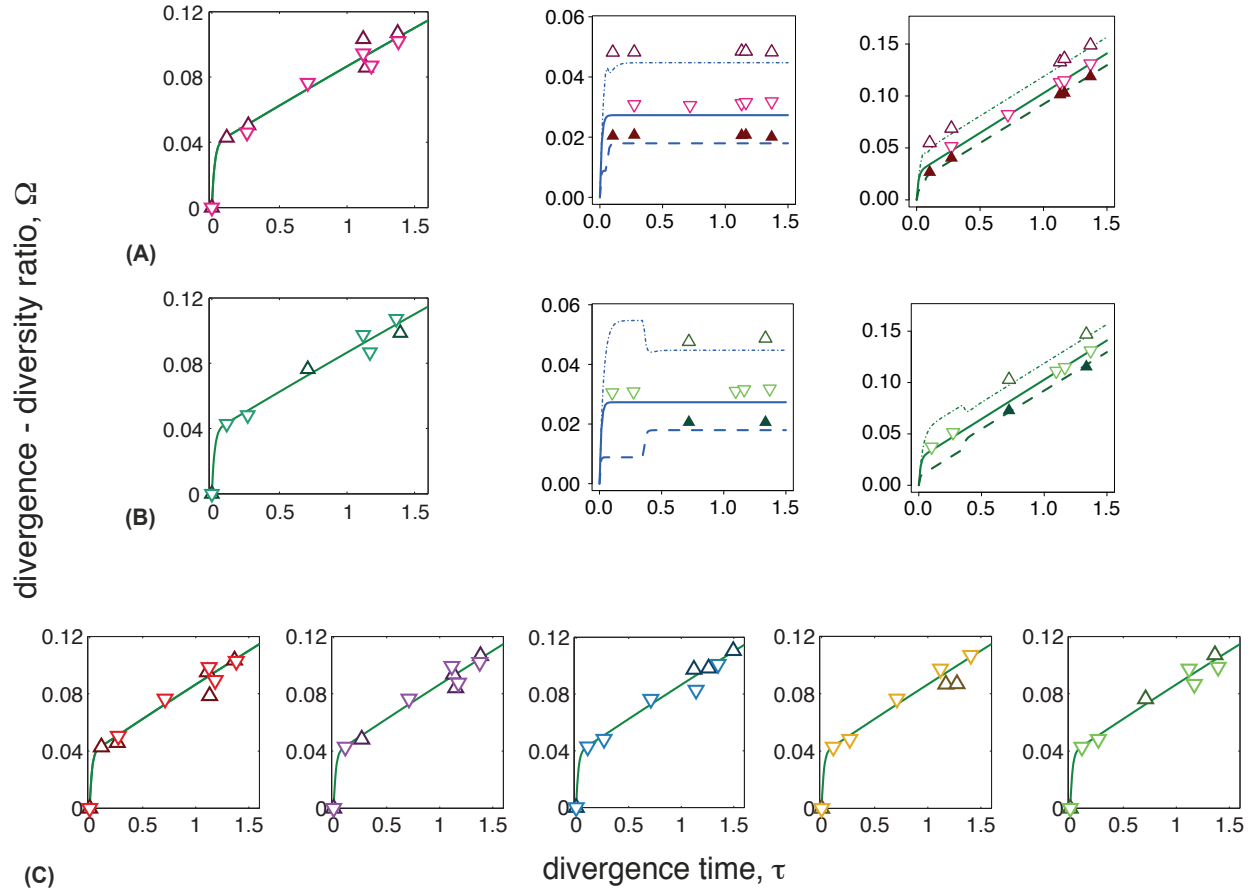


Figure S3: Test of lineage-specific demography. We compare the polarized divergence-diversity ratio $\Omega_{C,i}$ with species i as outgroup (equation 36, \triangle) to background data from partial clades *excluding* species i (∇); both quantities are plotted against the clade divergence time. **(A)** Left panel: Data for clades with outgroup *D. melanogaster*. Center and right panels: Evolution with a reduced or enhanced effective population size N_i in the outgroup lineage. Analytical curves and simulation results are shown for $N_i = 3N$ (dashed lines, \blacktriangle) $N_i = N/2$ (dashed-dotted lines, \triangle) in a fitness landscape ($c = 20, v = 0$; center panel) and seascape ($c = 20, v = 0.09$; right panel). **(B)** Same as (A), with outgroup *D. mojavensis*. **(C)** Data for each of the other five species chosen as outgroup. These data give no evidence of long-term lineage-specific demography. The analytical and simulation results show that lineage-specific demography under stabilizing selection does not give a spurious signal of adaptive evolution in the Ω test. Lineage-specific demography is introduced in section 3, simulation details are given in section 5 of Materials and Methods.

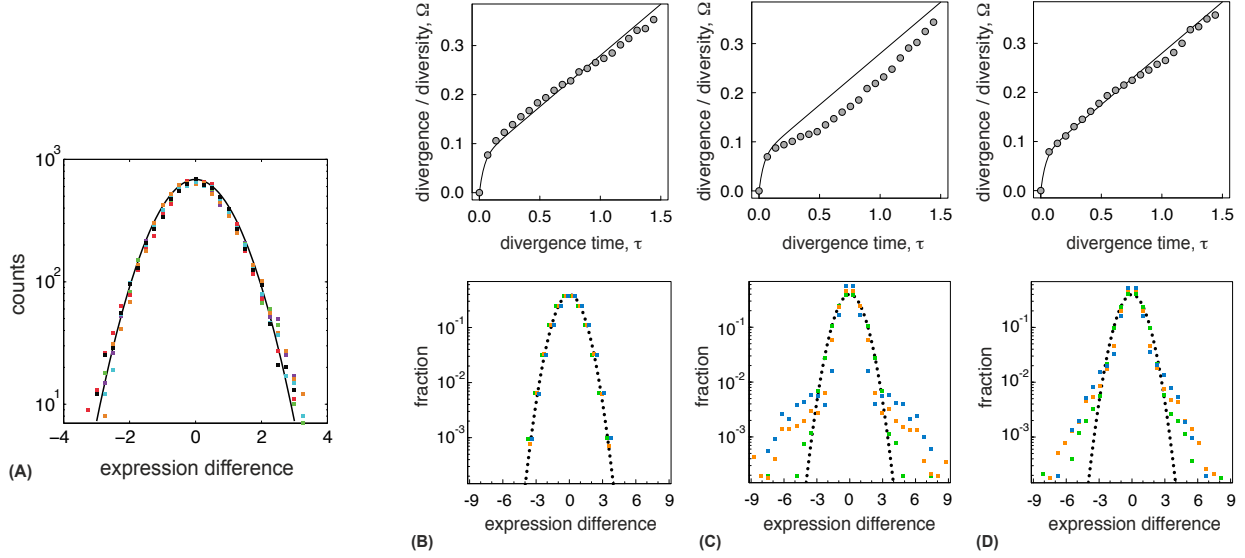


Figure S4: Test of alternative selection scenarios. (A) Distributions of clade-specific expression level differences, $P_C(\Delta E)$ (equation 39, color code as in Fig. S1A), standard-normalized to mean 0 and variance 1. These distributions are approximately Gaussian (black line: standard normal distribution). (B) Minimal seascape model. Top panel: Divergence-diversity ratio $\Omega(\tau)$ (bullets: simulation results; line: analytical curve as in Fig. 1A). Bottom panel: Standard-normalized distributions of trait differences, $P_\tau(\Delta E)$, from simulations for $\tau = 0.21, 0.69$ and 1.37 (green, orange, and blue bullets) are of Gaussian form (dotted line). The same quantities are shown for two alternative fitness models: (C) Loss-of-function model. Functional genes evolve in a static fitness landscape of stabilizing strength $c = 4.5$; individual genes lose function with rate $\gamma = 0.04 \mu$, resulting in reduced selection ($c \rightarrow 0.01 c$). The loss events generate a nonlinearity in $\Omega(\tau)$ and a broad tail in $P_\tau(\Delta E)$ that are not observed in the data. (D) Punctuated fitness seascape. Individual genes jump to a new, uncorrelated fitness peak with rate 0.16μ . These dynamics also generate a broad tail in $P_\tau(\Delta E)$. The *Drosophila* data of Ω_C (Fig. 1A) and of $P_C(\Delta E)$ together favor the minimal seascape model over both alternatives. The loss-of-function model and the punctuated seascape model are introduced in section 3, simulation details are given in section 5 of Materials and Methods.

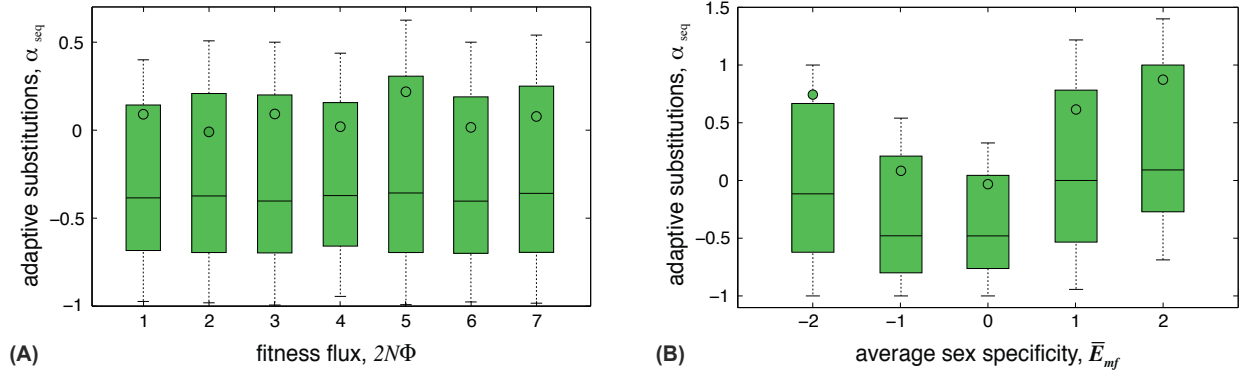


Figure S5: Adaptive gene expression versus adaptive evolution of protein sequence. (a) The distribution of $\alpha_{\text{seq}} = (D_n P_s / D_s P_n) - 1$, denoting the fraction of adaptive amino acid substitutions [57], is plotted against the cumulative fitness flux of gene expression, $2N\Phi$ (circle: average; line: median; box: 50% around median; bars: 70% around median). We find no correlation between these statistics, which suggests that adaptive gene expression is an independent mode of evolution. (B) The distribution of α_{seq} plotted against the average sex specificity \bar{E}_{mf} signals increased adaptive protein evolution in genes with sex-biased expression, which is strongest in male-biased genes (cf. the results of ref. [19]). For the definition of sex-biased expression, see section 4 of Materials and Methods.

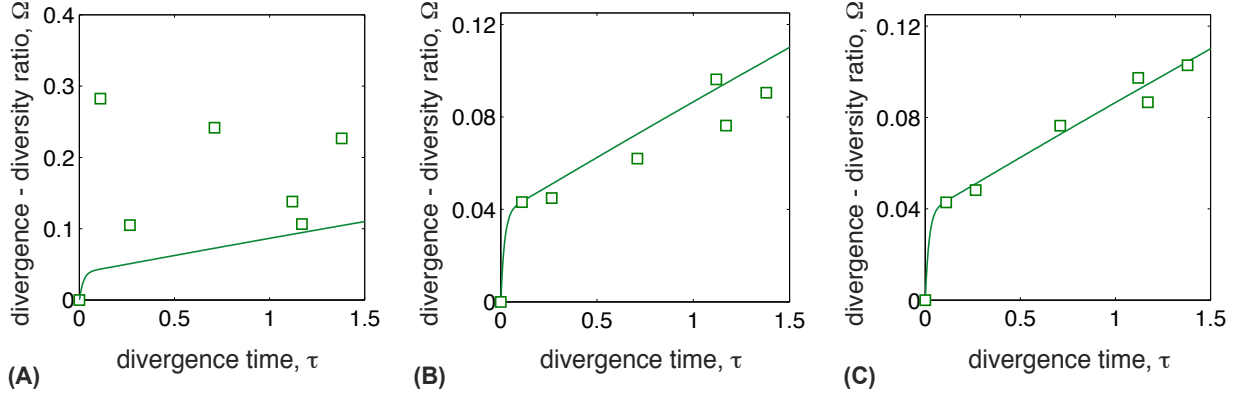


Figure S6: Rescaling of gene expression data. The clade-specific divergence-diversity statistics (Ω_c, τ_c) are shown for three variants of expression levels. **(A)** Raw data: \log_2 intensities averaged over experimental replicates. **(B)** Levels shifted to gene average 0 within each replicate line. **(C)** Levels shifted to gene average 0 and normalized to variance 1 within each replicate line [50], as used throughout this paper. The theoretical curve $\Omega(\tau)$ of the best-fit seascape model ($c^* = 18.6, 2N\Phi^* = 3.6$) is shown in all three panels. The additive shift in (B,C) is seen to be essential for evolutionary analysis, the subsequent multiplicative rescaling in (C) further reduces the noise in the Ω data. See section 1 of Materials and Methods.

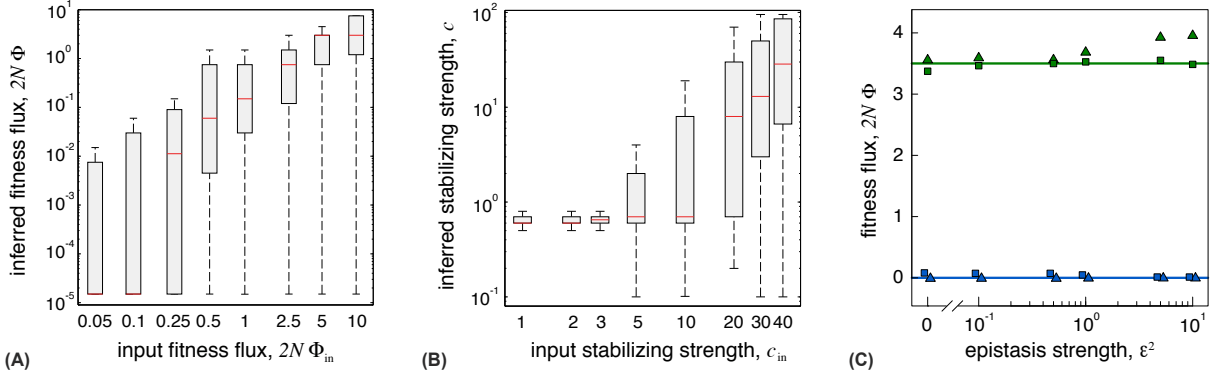


Figure S7: Simulation tests of the inference scheme. (A,B) Distributions of the cumulative fitness flux $2N\Phi^\alpha$ and stabilizing strength c^α inferred from simulated expression data are plotted against the simulation input parameters $2N\Phi_{in}$ and c_{in} (red line: median, box: 50% around the median, bar: 75% around median). See section 5 of Materials and Methods for simulation details. (C) Selection inference for epistatic traits. Simulation results of the actual fitness flux (\triangle) are compared to flux values inferred by the standard Ω test (\square , see section 2 of Methods). Both quantities are plotted against the strength of epistasis, ϵ^2 , defined as the ratio of epistatic and additive trait variance (section 5 of Methods); horizontal lines show the actual fitness flux without epistasis ($\epsilon^2 = 0$). Simulations are shown for selection parameters ($c = 4.5, v = 0.4$) (green) and ($c = 4.5, v = 0$) (blue). See section 5 of Materials and Methods for simulation details.

			mel-sim	mel-yak	vir-moj	mel-ana	mel-pse	mel-moj (Dros.)	
gene class (gene number)	c^*	$2N\Phi^*$	ω						α
all genes (6332)	18.6	3.6	8%	23%	47%	59%	60%	63%	54%
broad codon usage (1176)	15.6	3.9	9%	25%	49%	61%	62%	66%	57%
narrow codon usage (501)	19.0	2.3	5%	17%	36%	48%	49%	53%	18%
high expression (553)	14.3	1.7	1%	8%	27%	39%	40%	44%	0%

Table 1: Selection parameters and amount of adaptation. c^* : maximum-likelihood stabilizing strength. $2N\Phi^*$: maximum-likelihood fitness flux. ω : clade-dependent adaptive fraction of the gene expression divergence. α : fraction of adaptively regulated genes across the *Drosophila* genus, given by the condition $2N\Phi^\alpha > 4$. See equations (33) and (35) for definitions; gene classes are introduced in section 4 of Materials and Methods.